Determining Correlation Between Video Stimulus and Electrodermal Activity

Reza Tasooji

Thesis submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science in

Computer Science & Application

R. Benjamin Knapp, Chair
Denis Gračanin, Co-Chair
Tomas L. Martin

May 7, 2018
Blacksburg, Virginia

Keywords: Emotion Recognition, Affective Computing, Human Computer Interaction, Electrodermal Activity, Fear

Copyright 2018, Reza Tasooji
Determining Correlation Between Video Stimulus and Electrodermal Activity

Reza Tasooji

ABSTRACT

With the growth of wearable devices capable of measuring physiological signals, affective computing is becoming more popular than before that gradually will remove our cognitive approach. One of the physiological signals is the electrodermal activities (EDA) signal. We explore how video stimulus that might arouse fear affect the EDA signal. To better understand EDA signal, two different medians, a scene from a movie and a scene from a video game, were selected to arouse fear.

We conducted a user study with 20 participants and analyzed the differences between medians and proposed a method capable of detecting the highlights of the stimulus using only EDA signals. The study results show that there are no significant differences between two medians except that users are more engaged with the content of the video game. From gathered data, we propose a similarity measurement method for clustering different users based on how common they reacted to different highlights. The result shows for 300 seconds stimulus, using a window size of 10 seconds, our approach for detecting highlights of the stimulus has the precision of one for both medians, and F1 score of 0.96 and 0.74 for movie and video game respectively.
Determining Correlation Between Video Stimulus and Electrodermal Activity

Reza Tasooji

GENERAL AUDIENCE ABSTRACT

In this work, we explore different approaches to analyze and cluster EDA signals. Two different medians, a scene from a movie and a scene from a video game, were selected to arouse fear.

By conducting a user study with 20 participants, we analyzed the differences between two medians and proposed a method capable of detecting highlights of the video clip using only EDA signals. The result of the study, shows there are no significant differences between two medians except that users are more engaged to the content of the video game. From gathered data, we propose a similarity measurement method for clustering different user based on how common they reacted to different highlights.
Acknowledgments

I would like to express my sincere thanks to my advisor Dr. Benjamin Knapp for giving me the opportunity to explore this topic and for all of his invaluable feedbacks.

I would like to express my sincerest gratitude to my co-advisor Dr. Denis Gračanin for his continues support and guidelines. Without his help and advices, it was not possible to achieve the result for the current study.

I also feel very lucky to know and thankful to Dr. Deba Pratim Saha for all his assistance, feedback and helps through this study.

Finally, I would like to thanks all the participants who voluntarily, contributed in the study. Without them there was no data to even study.
# Contents

List of Figures viii

List of Tables x

1 Introduction 1

1.1 Motivation ......................................................... 2

1.2 Research Aims and Objectives .................................... 3

1.3 Document Organization ........................................... 5

2 Literature Review 6

2.1 Emotion and Affect Response ..................................... 6

2.2 Emotions .................................................................. 10

2.3 Studies in Fear ....................................................... 10
2.4      Physiological Signals for Measuring Emotion

2.4.1    Electrodermal Activity

2.4.2    Analyzing EDA Signal

2.5      Applications of Affective Computing

3      Problem Definition

4      Methods and Techniques

4.1      Assumptions and Limitations

4.2      System Architecture

4.3      Research Approach

4.3.1    Emotion Arousal

4.3.2    Experiment Design

4.3.2.1  Selecting Stimuli

4.3.2.2  Running The Experiment

4.3.2.3  Ethics Consideration

4.3.3    Data Collection

4.3.4    Data Processing
4.3.4.1 standardizing the data .............................. 37

4.3.4.2 Event Detection ...................................... 39

4.3.4.3 Similarity Measures and Clustering ..................... 46

5 Results and Discussion ................................. 51

5.1 Highlight Detection ..................................... 53

5.2 Similarity Measurement and Clustering ..................... 60

5.3 Discussion ............................................... 65

6 Conclusions and Future Works ......................... 68

Bibliography ................................................. 71
List of Figures

2.1  James Russell’s circumplex model for emotion mapping .......................... 8
2.2  EDA signal components. ................................................................. 18
4.1  System architecture. ................................................................. 27
4.2  Stimulus duration. ................................................................. 31
4.3  Post-questionnaires for the study. ............................................ 35
4.4  E4 wristband by Empatica. ...................................................... 36
4.5  Raw and standardized SC signal with phasic and tonic responses. ...... 39
4.6  Square sum and area under curve comparison. ............................... 40
4.7  Detected peaks from Z-score phasic response. ................................. 41
4.8  Detected peaks from Z-score phasic response. ................................. 42
4.9  Users appearing inside the bins. ................................................. 43
4.10 Output result. ................................................................. 45

4.11 Grouped individuals in event. ............................................. 49

5.1 The area under the curve and Square sum for EDA, tonic and phasic response. 54

5.2 Area under the curve of phasic response with detected peaks. ............... 55

5.3 The location of the peaks for all participants. ................................. 56

5.4 Histogram of peaks with different number of bins. .......................... 56

5.5 Histogram of peaks with different number of bins. .......................... 57

5.6 Grouped participants from different standardization methods. .............. 61

5.7 Grouped participants. .......................................................... 62

5.8 Clustered EDA signals. ....................................................... 64

5.9 Regression model for movie. .................................................. 67

5.10 Regression model for video game. .......................................... 67
List of Tables

4.1 Events in Movie Stimulus ........................................... 32
4.2 Events in video game stimulus. ................................. 32
4.3 Number of edges in each bin Level. ............................ 45
4.4 Groups in event. ...................................................... 49
5.1 Mean and Standard Deviation between movie and video game. .............. 52
5.2 Mean and Standard Deviation between movie and video game. .............. 52
5.3 Highlight detection including all signals for movie. ........................ 58
5.4 Highlight detection including all signals for video game. .................... 59
5.5 Highlight detection per individual. .................................. 59
5.6 Clustered participants using $k = 3$. ................................ 64
Chapter 1

Introduction

“Everyone knows what an emotion is, until asked to give a definition. Then, it seems, no one knows.”

- Beverley Fehr and James A. Russell, [18]

Our emotions play an important role in our decision making and the outcome of our decisions can change the emotions we experience. This interplay between our emotions and decisions plays a crucial role when defining normal and abnormal human behaviors.

In psychology, the term “affect” describes the experience of emotion. It is a sensory or behavioral reaction in our organism that is caused by any object or event. The object or event that causes such changes is called “stimulus”. The stimulus can be the cause of our emotion. But how can we use our emotion as input for applications?
1.1 Motivation

Imagine a system that correctly understands our emotion. Many researchers explored different methods and implementations to develop a system capable of understanding human emotions. Advancement in biometric sensors and computational powers, made researchers to become eager to implement an intelligent systems to gradually disappear our cognitive front.

According to UN’s Department of Economic and Social Affairs, Population Division, the number of older persons is growing faster than the number of people in all younger age groups and by 2050, people above 60 “are projected to account for one in five people globally” [51].

This will increase the need of Ambient Assisted Living (AAL) framework, an Context-Aware Intelligent Environments (CAIE) that is capable of assisting the users in their daily tasks including, providing e-Health solutions [38].

The output of CAIE has a great potential to be used in different areas such as entertainment industry or Ambient Assisted Living (AAL) frameworks. One of the aspects of CAIE is to automatically detect emotion. Automatic detection of emotions is one of the main goals of affective computing, also known is emotion AI [44].

One approach for detecting emotion is to use physiological signals. As Chenes mentioned in his study [8], the physiological signals, are difficult to hide or fake and with appropriate biometric sensors it can be possible to detect emotion.
For many years, medical and health services have been using biometric sensory devices to monitor the behavior of different organism such as heart, brain, muscle, etc, but recently these sensory devices are became more available to the public in different form such as wristband. For research purposes, this makes the process of collecting data easier than before which provides a better opportunity for analyzing physiological signals.

Exploring methods for analyzing and clustering different physiological signals not only can help us to better understand emotion but also can provide better e-health solutions, for example it will become possible to detect epileptic seizures, or level of stress or anxiety.

With the current advancements in this area, this complex system, yet, receives limited attention in empirical research.

1.2 Research Aims and Objectives

Our goal is to explore methods for measuring the correlation and causation between physiological signal and video stimulus.

Our hypothesis is:

*The EDA signal and the features that can be extracted from the EDA signal might be sufficient for detecting affect arousal highlights caused by different stimulus.*

To test our hypothesis, we proceed in two stages. First, collecting the data which requires conducting the user study and providing the infrastructure to receive and store physiological
signals. second, after gathering the signals, analyze the signal and develop an algorithm to find out different highlight of the stimulus from EDA signals.

For the first stage of our research, it is important to provide consistency between users’ emotion and to have control over the cause of arousal in the stimulus. Consequently, we decided to focus on content that might arouse fear.

Two different median were used in the study as stimulus. That allows us to compare different level of arousal between users and to see if there are any significant differences between different stimulus.

For one stimulus we decided to show the section of the “Halloween” movie and for another stimulus, we decided to show the section of the video game named “Evil with it”.

In the second stage of our research, we analyze EDA signal and explore methods for detecting the highlight of the stimulus, using only EDA signals. Beside this, we present a method for measuring similarities between participants based on the collected EDA signals.

The results can be used in future works to make an application capable of clustering users and be able to decide, based on the physiological changes of the user, what to show from different content of the stimulus.
1.3 Document Organization

In Chapter 2, we explain what emotion is and cover some of the related works in field of emotion recognition. We cover some studies related to fear and how to analyze EDA signal, we finish the chapter by covering some of the studies that they used physiological signals to build different forms of applications, such as detecting stress or anxiety.

In Chapter 3 we define our hypothesis and what separates this study from other studies in this field.

In Chapter 4, we cover our methods and approach for testing our hypothesis in more detail. We start with conducting the user study and then describe our approach for collecting the data and the steps to process our data. We cover our approach for detecting highlights of the stimulus based on the EDA signals. Then we propose a method for measuring the similarities between participants and how to use this information to cluster the users.

Chapter 5, we explore the results from our study. We start with analyzing self-report questionnaires from the study and after that we explore the result from event detector followed by the results from clustering users.

Chapter 6 discusses our finding and proposes some possible future works that can be done to improve the result from this study.
Chapter 2

Literature Review

This chapter covers some of the related works in the field of physiological signals and emotion recognition. First we explain what emotion is, and how it can be recognized by using physiological signals. Then we explore some of the studies about fear. Then, we focus on EDA signals as a data for analyzing physiological changes. We explore in detail what EDA signal is and how to analyze the signal. Finally, we look at some of the papers that implemented systems capable of using physiological signals for various reasons.

2.1 Emotion and Affect Response

According to the Diagnostic and Statistical Manual of Mental Disorders-IV-TR (American Psychiatric Association, 2000, p. 819) affect is a “pattern of observable behaviors that is the expression of a subjectively experienced feeling state (emotion)”. Jaak Panksepp [42]
stated: “Emotion is the umbrella term for all of the behavioral, expressive, cognitive and physiological changes that occur. Affect is the conscious experience of an emotion”.

Emotion is the state of our feeling and affect is the appearance of our emotion. Thus an affect is the observable action caused by the emotional state. But emotions are complex and by just observing the affect it is not possible to completely recognize the right emotion without uncertainty.

In the history of emotion study, some researchers suppose emotions to be essentially universal. In this approach, what can be observed are what is known as "basic emotions". The core concept of basic emotions is based on discrete emotions theory. This theory claims that there is small number of core emotions that can be distinguished by an individual’s facial expression and biological processes that are universal among all people [1, 17]. Different studies group basic emotions to different numbers. [14, 25, 42, 49]

The most commonly classification for basic emotions are these seven emotional states: Happiness, sadness, anger, fear, disgust, surprise and stress. During the past decades lots of studies used these concepts to find the accurate process to link the right affect to the right emotion to answer this question that is it possible to build a system capable of predicting the correct emotion based on the affect?

The result form these studies mainly focus on analyzing individual emotion elements such as facial expressions. For example, James Russell claims some emotions are universally recognized from facial expression” [1]. To support his claim he carried out his study in
western cultures and cultures isolated from western influences. The result of this study shows that there is no significant effect of culture for “happy”, “surprise” and “sadness” expression, but there exist significant effect of culture in “disgust”, “anger” and “fear”. Ekman [13] believes because the same facial muscular movement is associated with the same emotions in all people, it is fundamentally genetically determined and it is inherited.

Another group of researchers believe cultural differences exist in some aspect of emotions and emotion can not only be determined biologically [1, 34, 36]. Environment, social or cultural situations also determine emotions.

To better understand the differences in emotion between cultures, first we need to explore theoretical methods for mapping emotion. In 1980, psychologist James Russell [47] used the Cartesian space to visualize the connection between arousal and emotions (Figure 2.1). In his paper [47], he proposed the *circumplex model of affect*. The circumplex model of affects proposes that all emotions can be structured in two fundamental dimensions; valance and arousal.
In his model, the horizontal dimension represents pleasure-displeasure in affective state, also known as valance. The vertical dimension is degree of arousal which also known as activation-deactivation or engagement-disengagement.

For example, a relaxed person is located on the right side of the model (pleasure) without any value on arousal axis or an angry person is located on left side of the model with high value on arousal axis. The center point in this model is a neutral mood.

The study by Feldman [19], shows that the degree of arousal can be related to the amount of physiological activities during affective experience. In her study she shows in self-reported moods, people weight the arousal dimensions less than the valence dimensions, while in the semantic circumplex structure [47], the two dimensions are weighted equally.

Different studies show that different cultures have different level of arousal [32, 36]. The finding from the study by Matsumoto [36] shows, Japanese make lower intensity rating to the posers that represent emotions, than the Americans and different expressions has different rating along different cultures.

The study by Lim [32] shows that in Eastern cultures, low arousal emotions (sadness, relaxed) are valued more than high arousal emotions (happiness, angry), while in western cultures, high arousal emotions are valued and promoted more.
2.2 Emotions

There are different models to theoretically classify emotions. The “Wheel of Emotions” by Robert Plutchik [45] categorizes emotion into eight primary emotions known as, joy, acceptance, fear, surprise, sadness, disgust, anger and anticipation. In his model, every primary emotion comes with bipolar emotion. For example, joy versus sadness, anger verses fear, surprise vs anticipation, etc.

The circumflex model by James Russell’s (Figure 2.1) is another commonly used model for categorizing different emotions. He used two dimensions, arousal and valence in his model to categorize emotion. The valence dimension classify emotion based on being negative (unpleasant) or positive (pleasure). Arousal is how weak or strong our nerve system react to stimulus. This opens possibilities to introduce our emotion for affective computing.

In this study, we are going to mainly focus on arousal dimension in the circumflex model.

For this approach we need to select the location of emotion in valence dimension. For this reason, we are going to mainly focus on fear and study the level of arousal caused by stimulus.

2.3 Studies in Fear

“The human mind is fascinated with the unknown and to most, the unknown is terrifying.” [41]

In our study, we mainly focused on stimulus that might arouse fear. Here we mention some
of the studies that explore what fear is and what might cause fear.

Walter Cannon describe fear as an evolutionary necessity that notify a person to proceed on their current direction or take another action in order to increase the likelihood of survival. The excessive fright can cause a cognitive dissonance state, which can cause an attitude change or an inconsistent thoughts that might leads anxieties into phobias.

Fear is an essential emotional response to societal conflicts. For example, fearful expression is thought to serve as a social cue that a person with this expressions, might avoid antisocial behaviors [35].

In the study by Adam Palmer [41], he looked at the individual differences in people’s behavior to horror film. The result of his study shows Systolic Blood Pressure (SBP) was the only physiological signal that increased significantly from the baseline to the film. He also claims there is physiological differences between people that watch horror movies and people that do not. In his research, he found that fear, in particular, can cause physiological changes such as face temperature increase, higher skin conductance activities and increase in muscle potential and respiration rates.

In the study by Cuthbert et al. [40], the result shows the phobic patients showing higher level of arousal while the panic disorder patients showed the least.

In the study by Teresa Lynch [33], she examined fright experiences caused by video games. The result shows different elements such as darkness, zombies, being surprised, etc. in video games are causes of fear. It also shows there is no significant differences between male and
female on experiencing fear. This study is based on the self-report on 269 undergraduate students and no physiological signal measurement were included in this study.

A. Dasgupta et al. [10] developed a proof of concept Mixed Reality application that uses exposure therapy treatment for social anxiety disorder. In her study, she recorded different physiological changes such as heart rate, inter beat intervals, temperature and electrodermal activity for analyzing her study. The result from the pilot study shows the effectiveness of the MR approach for classifying psycho pathologies and understanding behavioral patterns.

The result from the study by Trubanova et al. [50] shows the positive correlation between recognizing fear and perspective-talking abilities, while perspective-taking was not significantly associated with recognition of the other basic emotions. Perspective-taking refers to the ability to perceive, appreciate and consider the perspective, or point of view, of another individual. The result of the study shows there is something unique about fear that might separate it from other basic emotions.

2.4 Physiological Signals for Measuring Emotion

Although the area of speech recognition and facial expression for identifying emotion are far more explored [9,12,15,20], other lesser-used methods are the use of bio-sensors to measure and analyze physiological signals.

Different studies show that, the two fundamental dimensions of emotion, valence and arousal, are related to physiological aspects and brain activities [23,29,39].
The physiologic part of the emotion is closely connected to arousal of the nervous systems. The nervous system is divided into two sections, Central Nervous System (CNS) and the Peripheral Nervous System (PNS). The PNS is also divided into somatic and visceral (autonomic nervous system) parts. The autonomic nervous system (ANS), also known as vegetative nervous system, is responsible to control internal organs that are not consciously directed, such as the heartbeat or breathing. It regulates the body’s unconscious actions.

ANS is divided into two components: the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS) [26].

SNS primary process, is to simulate fight-or-flight response, which is “physiological reaction that occurs in response to a perceived harmful event, attack, or threat to survival” [7].

ANS is connected to different organism, such as smooth muscles, cardiac muscles and the glands of internal organs, that produce slight electric pulses. By measuring these electric pulses during affect response, it is possible to measure the level of arousal and recognize emotion.

Although these electric changes alone does not identify which specific emotion is being elicited but one common way to measure emotion is to analyze the changes that happens in autonomic nervous system [3, 6]. For example, changes in heart beat, blood pressure, galvanic skin responses, also known as electrodermal activity, or the enlargement of pupil diameter under fear stimuli are ways to recognize emotion based on ANS.

Ekman [16] manually analyzed different physiological signals such as finger temperature,
heart rate and skin conductance, that occurred during stimulus that might cause anger, fear, disgust, and happiness. The result of his study shows, sadness, fear and anger cause a larger increase in heart rate than disgust, while anger produced a larger increase in finger temperature than fear. Also both fear and disgust produced larger skin conductance than happiness.

Gross and Levenson [22] studied three emotions: amusement, neutrality and sadness by using films as their stimuli. 180 female, participated in their study to watch sad, neutral and amusing films. Skin conductance, inter-beat interval, pulse transit times and respiratory activation were measured. Their results shows the increase in inter-beat interval for all three states, while neutrality is less than amusement and sadness. Skin conductance increased after the amusement film, decreased after the neutral film and stayed the same after the sadness film.

From these studies we can conclude that physiological signals can be a reliable data for detecting the level of arousal, although it might not be possible to detect the basic emotion with high accuracy, without knowing the content of the stimulus.

### 2.4.1 Electrodermal Activity

In this study we using only EDA signals for measuring arousal. This section describes in detail what EDA signal is and some methods for analyzing EDA signals.

The term EDA introduced by Johnson and Lubin in 1966 as the umbrella term for all
electrical phenomena related to skin \[27\]. Since 1849 when Electrodermal activity where first observed by DuBois-Reymone in Germany, Electrodermal recording has become one the most frequency used bio-signals in psychophysiology.

EDA is the term which defines the changes in the electrical properties of the skin “due to the activity of sweat glands”, this can be altered by sweating or blood flow during sympathetic nervous system (SNS) activities, across specific active sites.

There is about three million sweat glands in human body and the greatest density of sweat glands can be found on the palms, soles which makes these sites as active sites for measuring EDA \[11\].

Electrodermal recording is usually performed with two electrodes, by placing them on two active sites such as palm. “The most common electrode placements are the thenar eminences of the palms and the volar surface of the medial or distal phalanges of the fingers” \[11\].

The quantified EDA signal composed of two components, Electrodermal Level (EDL) and Electrodermal Response or reaction (EDR).

The EDL, also known as tonic is the slow background changes. The EDL response can be describe as slow changes that happens in absence of stimulus or during response-free recording intervals.

On the other hand, using the term response or reaction in EDR, suggests there is a distinct relationship to stimulus that might cause EDR.

These two components of EDA signal, makes EDA signals a very reliable indicator of SNS
activity and the main reason for the popularity of EDA signal is the ease of obtaining Electrodermal Responses (EDR) [6]. This parameter of EDA signal can be used to detect highlight of the stimulus.

Some other parameters that can be obtained from EDA signals are as follow: From the phasic response we can gather, frequency (SCR freq) which is number of EDR in a given time window, amplitude (SCR amp), which is the height of the single response, latency (SCR lat) and rise time (SCR rise.t). Figure 2.2 shows EDA signal and some of the parameters that can be extracted from EDA signal. The figure shows what can be extracted as phasic activity (SCR). It shows a steep incline to the peak known as rise time and slow decline to the baseline.

In next section we cover methods for analyzing EDA signal to gather some of the parameters that we mentioned.

2.4.2 Analyzing EDA Signal

One of the classic approaches to analyze EDA signal is to use the standard peak detection method (trough-to-peak) [6]. In this method, SCR amplitude is calculated by measuring the differences of the EDA values at its peak and at the preceding trough.

One problem with this approach is how to handle and detect SCRs that are closely superposed.

To solve this problem, Benedek and Kaernbach [4] proposed a method known as “continuous
decomposition analysis” to deconvolve EDA signal into two contentious signals of tonic and physic activity.

This method is based on the precondition that there exist a stable Impulse Response Function (IRF). IRF is an assumption that sudomotor nerve activity will show peaks, known as sudomotor bursts, with short time constants which leads to a larger time constant exhibition in SCRs. The IRF represents the basic SCR shape that would result from a unit impulse. a Biexponential function with the parameter of $\tau = 0.75s$ and $\tau = 2s$ was found to represent the IRF adequate for deconvolution process [2].

$$IRF = c.(e^{-\frac{t}{0.75}} - e^{-\frac{t}{2}})$$

Another approach that might result to the basic SCR shape, is to pop a balloon. This method might cause a unit impulse that can be used as IRF, without modeling the shape. The deviation of the data from a model IRF result in driver responses, which might reflect the activity of the sudomotor nerve. The sudomotor nerve activity is the cause of sweat secretion and triggers any changes in skin conductivity, thus it can be considered as a driver, consisted of a sequences of distinct able impulses which triggers specific impulse response (SCRs).

With the precondition that there exist a stable IRF, SC activities can be calculated as follow:

$$SC_{\text{Phasic}} = Driver_{\text{Phasic}} \ast IRF$$
The same equation is true for tonic activities. EDA data can then represented as:

\[ SC = (Driver_{\text{Tonic}} + Driver_{\text{Phasic}}) \ast IRF \]

\[ \frac{SC}{IRF} = (Driver_{\text{Tonic}} + Driver_{\text{Phasic}}) = Driver_{SC} \]

In the absence of any phasic activities we can observe tonic activities.

With this approach it become possible to deconvolve the EDA signal into phasic and tonic responses.
2.5 Applications of Affective Computing

Jennifer Healey et al. conducted a study [24] to detect the stress level of the drivers by measuring different physiological signals (electrocardiogram, electromyogram, respiration, and skin conductance.) In their study, they used SFFS (Sequential Forward Floating Selection) algorithms to recognize pattern of drivers’ stress. The result of their study shows, they were capable of recognizing the intensity of driver’s stress by 88.6%.

D. P. Saha, et al. [48], explores the possibility of determining the relevance of services provided by an intelligent environment by creating an affective feedback loop. The author tries to create an infrastructure-level fusion between affective computing and context-aware systems, to provide a communication between a user and intelligent environment. In their system the communication between user and environment were caused by sensing stress. The result from this study shows that by training the Space Vector Machine (SVM) models individually for each user, it might be possible to find the similarity in the patterns of physiological data. The result shows it is possible to use affective computing for developing context-aware applications.

Bornoiu [5] presents a method for identifying the stress level by using a Kohonen neural network. Different parameters were extracted from the EDA signals: rise time and amplitude of EDA signal, Skin Conductance Level (SCL) gradient, Skin conductance Response (SCR) power and frequency. Their stimulus for arousing stress has three parts. The first and the last part are to relax the user by playing classic music while slowly panning panoramic image
for two minute. The second part, the participate needs to solve mathematic equations with a variable time for each equation to cause stress.

Their implementation for detecting stress in the signal has an average recognition rate of 86.25% for their test set.
Chapter 3

Problem Definition

The exponential growth of video media, demands, more than ever before, to have some algorithms to summarize the content of the video. Personalized recommendations from content providers such as Netflix’s “Top picks for you” help both consumers to watch the content that they care about, and advertisement companies to provide them with better insight on what gets watched, and by whom.

Beside this, Virtual Reality (VR) and Augmented Reality (AR) are finding their ways into entertainment industry.

The entertainment industry is growing more than ever and with many variation on this median, the consumer want the content to be their way.

One of the approaches to put the consumer first in selecting the content, is to build a system capable of automatically detects what consumer want to see based on their physiological
Detecting the highlight of the media and analyzing the content, can help us to take the first step for building such system.

The process of automatically, detecting the highlights of video is a far-reaching subject. With the huge variation in genres and different median such as video game, movies, VR and AR, defining what needs to be consider as highlight can be diverging.

Designing a system capable of detecting such highlights comes with some limitations and it will not be able to handle all possible outcomes.

Some of the studies for detecting the highlights of the video clips are as follow:

Joho et al. [28], analyzed facial activities to detect personal highlights of the content. They developed real time facial expression recognition system that outputs a vector of motion features of certain regions of the face. In his study, there are total of 12 Motion Units (MU) on face that are being used for classification. They used 8 video clip from different genres with different durations and showed the content to 10 participants. To extract the highlight the author explore high level of consensus on a personal highlights for different video. In his study only one video has a common highlight among all users except one.

Chênes et al. [8], presents a technique to obtain user-independent summery of a given video. His approach does not require emotion recognition. The system concept is based on the physiological linkage between different participants’ emotional responses. four physiological signals were used in the study, Electromyogram (EMG) which measures the activities of the
muscles, Blood Volume Pulse (BVP), measures the change of blood pressure, Electrodermal activity (EDA) and skin temperature. The result of this study shows that skin temperature with a window response of 8 second returned the best correct classification rate (77%).

These studies show that it is possible to detect highlights of the stimulus, using physiological changes with high accuracy.

One of the problems with current studies for detecting highlight is that the process for collecting signals often laborious and exhausting for ordinary users to try them in their daily tasks. It might requires camera or a variation of biometric sensor that makes the process complex.

To solve this problem, in our infrastructure, we are using a wristband capable of recording different physiological signals. We believe this infrastructure can make the process of collecting physiological data much easier.

Beside this, most of the studies in the field of affective computing, use different physiological signals in their study. Although this approach will increase the accuracy of the results but it requires more processing and it might not possible to analyze the data in real time.

Our hypothesis is EDA signals has the essential parameters such as tonic and phasic responses that are sufficient for detecting highlights of the stimulus.

To test our hypothesis we conduct a user study with two stimulus that might arouse fear. The post questionnaire is being used to analyze the differences between two stimulus.

After that we propose a method for measuring the similarities between participants and
cluster them.

We plan to use the proposed method for detecting highlight and clustering in a real time application to build a system that automatically selects the best highlight of the stimulus based on the EDA signal of the users.

We hope results from this study help other researcher to implement affective computing system with higher accuracy that can be used by larger group of participants.
Chapter 4

Methods and Techniques

In this chapter, we discuss our assumptions for implementing our affective computing system. We describe system architecture and what different components are required for such systems. We explore our method for collecting data and how we developed our stimulus for collecting data. Then, methods that we used for processing the data are discussed. Here we cover methods for standardizing the data, measuring similarity between signals and finally we cover how the system detect highlights of the video clips.

4.1 Assumptions and Limitations

For our study, we mainly focused on analyzing EDA signals without including other physiological features such as heart rate or facial expressions. Our goal is to see if it is possible to build a system capable of clustering users and detecting highlights of the stimulus using
just EDA signals.

For this purpose, we assumed recording and analyzing EDA signals by itself, without including other source of physiological signals such as ECG (Electrocardiogram) or EEG (Electroencephalogram) might be sufficient for extracting required features for the system.

Our assumption comes with the idea that tonic and phasic responses are two main components of EDA signals that can be extracted from the main signal and phasic change in EDA signals play an important role on detecting highlights in stimulus.

The result of the study by Chênes et al. [8], shows that the EDA signal can be considered as the most promising signal for detecting highlights in video clips.

To reduce the level of ambiguity in the system, we decided to focus on an stimulus that might arouse fear.

The highlights of the stimulus are predefined by the author. We did not include self-report highlights by users in the current study to have consistency among the output for all data.

In our stimulus, any visual or auditory changes such as, change in the scene or sound, element of surprise in the stimulus, possible contents that might cause anticipations, etc, were marked as highlights.

The proposed method for detecting highlight of the stimulus has high precision and F1 score when the system uses signals from a group of users, and low precision and F1 score when the system uses individual signals.
4.2 System Architecture

Ray [46], presented an architectural framework for Home Health Hub Internet of Things for monitoring health of elderly people at home. The architecture of our system shares some similarities with that framework [46].

For our purpose we divided the system into two section. The first section, includes conducting user study. This includes the infrastructure for collecting and synchronizing the collected data. The output of this part of the system goes to the “Analyzer” layer. The “Analyzer” is the part of the system that we will test our hypothesis. Figure 4.1 shows different aspects of our system. The left part of the figure shows the process for collecting the data and the right part show how we process the signal for testing hypothesis.

The purpose of the infrastructure is to collect the data and make sure the collected data is synchronized to the highlights of the stimulus. Highlights are any event in the stimulus that might cause physiological changes. Sudden change in sound volume, change of scene or any
content that might surprise the audience can be considered as highlight. By knowing the highlights and the time that they appear, it become possible to analyze physiological signals caused by different events in the stimulus. Different biometric sensors can be used to stream and, or record physiological activities.

Before sending the data to the processing unit, the data needs to be synced with the stimulus. For this reason the physiological data and the information about stimulus needs to be transferred to a local machine to be synchronized. The synchronized data then will be sent to the analyzer.

The analyzer plays a vital role for processing the synchronized data. The purpose of the analyzer is to implement methods for testing the hypothesis.

To test our hypothesis that, using only EDA signal can detect highlights of the stimulus, we develop a method that uses different parameters from EDA signals for detecting highlights of the stimulus. Beside this we propose a distance base clustering method that group users based on how common they react to different highlights.

Finally, it might become possible to use the output from these two methods to implement an affective computing application capable of recommending the next highlight of the stimulus based on the current physiological changes of the user.
4.3 Research Approach

This section covers our approach for implementing different parts of the system. We start with our approach for building the stimulus for the study and how we conducted our user study. Then, we cover methods and infrastructure for collecting data. Finally, we cover methods for analyzing the data.

4.3.1 Emotion Arousal

The first step before collecting data is to provide the stimulus that might arouse emotion. In the review article about how to measure emotion by Mauss [37], the author concludes “there is no gold standard measure of emotional responding.” also later on she mentions that “emotions are constituted by multiple, situationally and individually variable processes.”

Human emotion is a complex process with multiple variables. To reduce the number of variables for our research, we conduct a user study to give us more control over different variables. In next section, we cover our approach for conducting our user study.

4.3.2 Experiment Design

For collecting EDA signal, we designed an experiment that only focuses on elements that might arouse fear. This decision was made to provide consistency among all samples. Beside this, the goal of the study is to analyze physiological signals caused by well known stimulus
to gives us a better insight about the characteristic of the signal.

Below we describe in detail our approach for selecting different stimulus for our study.

4.3.2.1 Selecting Stimuli

To provide different level of arousal, we selected two different stimulus for the study. The first stimulus is a section of the “Halloween ” movie (1981). The result of the study by Philippot [43] shows that the selected section of this movie might arouse fear. The second stimulus is a section of the video game named “Evil Within” (2014).

Our hypothesis is the content of the video game might have higher arousal level, compare to the “Halloween” movie. We assumed, because the content of the movie is old, it might lead to lower arousal level, compare to the video game with newer content.

Each stimulus includes 3 sections (Figure 4.2). The first section (30 seconds) named as “Baseline” is a black screen without any sound, is to collect the baseline of the user. The second section (90 seconds) called “Relaxed”, and the purpose is to relax the user. The content for this section were selected from the video game named “Flower” (2009). This section is similar for both movie and video game stimulus. Finally, “Fear section” are “Halloween” movie for movie stimulus and “Evil Within” for video game stimulus (210 seconds).

To be able to analyzing the EDA signal, time of different events for each stimulus were marked by the author.

Tables 4.1 and 4.2 show the name and the time of the events for movie and video game
Figure 4.2: Stimulus duration.

stimulus. The time is in seconds. It is the duration between the starting point of the stimulus and the moment the event happens in the stimulus.
Table 4.1: Events in Movie Stimulus

<table>
<thead>
<tr>
<th>Event Name</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start Relax Sound</td>
<td>32.814</td>
</tr>
<tr>
<td>Start Flower Scene</td>
<td>43.092</td>
</tr>
<tr>
<td>Start Wind Sound</td>
<td>56.132</td>
</tr>
<tr>
<td>Flower Ding Sound</td>
<td>75.851</td>
</tr>
<tr>
<td>Start Hum Sound</td>
<td>86.331</td>
</tr>
<tr>
<td>Start Strong Wind Sound</td>
<td>97.492</td>
</tr>
<tr>
<td>Cloud Scene</td>
<td>110.091</td>
</tr>
<tr>
<td>End Relax Scene</td>
<td>115.252</td>
</tr>
<tr>
<td>Start Movie Scene</td>
<td>120.091</td>
</tr>
<tr>
<td>Stairs Scene Change</td>
<td>135.571</td>
</tr>
<tr>
<td>Person On Stair Scene Change</td>
<td>141.211</td>
</tr>
<tr>
<td>Stair FPS Scene Change</td>
<td>147.132</td>
</tr>
<tr>
<td>On Sec Floor Scene Change</td>
<td>154.772</td>
</tr>
<tr>
<td>Show Door Scene Change</td>
<td>163.011</td>
</tr>
<tr>
<td>Person Scene Change</td>
<td>171.612</td>
</tr>
<tr>
<td>Show Victim Face</td>
<td>185.011</td>
</tr>
<tr>
<td>Open Door Scene</td>
<td>190.051</td>
</tr>
<tr>
<td>Start Loud Noise</td>
<td>199.052</td>
</tr>
<tr>
<td>Victim Shocked Scene</td>
<td>213.892</td>
</tr>
<tr>
<td>Show First Dead Person</td>
<td>224.252</td>
</tr>
<tr>
<td>Show Second Dead Person</td>
<td>234.931</td>
</tr>
<tr>
<td>Show Killer</td>
<td>255.492</td>
</tr>
<tr>
<td>Killer Hit Victim</td>
<td>260.411</td>
</tr>
<tr>
<td>Victim Falls From Second Floor</td>
<td>265.091</td>
</tr>
<tr>
<td>Theme Song Changed</td>
<td>271.011</td>
</tr>
<tr>
<td>Victim Running</td>
<td>277.571</td>
</tr>
<tr>
<td>Start Ding Ding Sound</td>
<td>291.092</td>
</tr>
<tr>
<td>Killer Breaks door</td>
<td>308.732</td>
</tr>
<tr>
<td>Killer Opens Door</td>
<td>318.291</td>
</tr>
<tr>
<td>Victim Breaks Glass</td>
<td>323.772</td>
</tr>
<tr>
<td>End Movie</td>
<td>328.572</td>
</tr>
</tbody>
</table>

Table 4.2: Events in video game stimulus.

<table>
<thead>
<tr>
<th>Event Name</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start Relax Sound</td>
<td>30.676</td>
</tr>
<tr>
<td>Start Flower Scene</td>
<td>43.635</td>
</tr>
<tr>
<td>Start Wind Sound</td>
<td>55.475</td>
</tr>
<tr>
<td>Start Ding Sound</td>
<td>75.155</td>
</tr>
<tr>
<td>Start Hum Sound</td>
<td>81.915</td>
</tr>
<tr>
<td>Start Strong Wind Sound</td>
<td>97.156</td>
</tr>
<tr>
<td>Cloud Scene</td>
<td>105.076</td>
</tr>
<tr>
<td>End Relax Scene</td>
<td>114.436</td>
</tr>
<tr>
<td>Start Music</td>
<td>120.715</td>
</tr>
<tr>
<td>Start Game Scene</td>
<td>127.035</td>
</tr>
<tr>
<td>Looks at hand</td>
<td>131.916</td>
</tr>
<tr>
<td>Looks at Body Upside Down</td>
<td>148.955</td>
</tr>
<tr>
<td>Killer Introduced</td>
<td>151.155</td>
</tr>
<tr>
<td>Cuttin Sound</td>
<td>163.236</td>
</tr>
<tr>
<td>Loud Cutting Sound</td>
<td>171.876</td>
</tr>
<tr>
<td>Killer Graps Dead Body</td>
<td>186.596</td>
</tr>
<tr>
<td>Found Knife</td>
<td>196.955</td>
</tr>
<tr>
<td>Start Swinging</td>
<td>199.555</td>
</tr>
<tr>
<td>Second Swing</td>
<td>202.875</td>
</tr>
<tr>
<td>Third Swing</td>
<td>206.235</td>
</tr>
<tr>
<td>Try Catch Knife</td>
<td>209.356</td>
</tr>
<tr>
<td>Second Try Catch Knife</td>
<td>212.435</td>
</tr>
<tr>
<td>Third Try Catch knife</td>
<td>215.836</td>
</tr>
<tr>
<td>Caught Knife</td>
<td>219.195</td>
</tr>
<tr>
<td>Relased From Rope</td>
<td>223.795</td>
</tr>
<tr>
<td>walking To Killer</td>
<td>243.756</td>
</tr>
<tr>
<td>UI Displayed</td>
<td>244.556</td>
</tr>
<tr>
<td>Hiding Behind Wall</td>
<td>250.275</td>
</tr>
<tr>
<td>Gets The Key</td>
<td>259.076</td>
</tr>
<tr>
<td>Chainsaw Sound</td>
<td>324.236</td>
</tr>
<tr>
<td>Opens Door</td>
<td>326.756</td>
</tr>
<tr>
<td>Cut The Leg</td>
<td>338.636</td>
</tr>
<tr>
<td>Door Opens</td>
<td>342.235</td>
</tr>
<tr>
<td>Looking Down</td>
<td>310.635</td>
</tr>
<tr>
<td>Wire Triggered</td>
<td>318.236</td>
</tr>
<tr>
<td>Chainsaw Sound</td>
<td>324.236</td>
</tr>
<tr>
<td>Opens Door</td>
<td>326.756</td>
</tr>
<tr>
<td>Cut The Leg</td>
<td>338.636</td>
</tr>
<tr>
<td>End Game</td>
<td>342.235</td>
</tr>
</tbody>
</table>
For our study, to be able to compare two median, movie and video game, we decided to remove any element of interaction between user and the content by asking participants to watch prerecorded section of the video game instead of asking them to play the content. This helps to synchronize the time of the events with the time of the recorded EDA signal and compare the result to movie stimulus.

4.3.2.2 Running The Experiment

For each participant, we divided the study into two sections at least one week apart. In first section, we asked participants to watch the stimulus that contains the scene from the movie and for the next section we ask them to watch the clip with video game content while their physiological signal being recorded.

To reduce the element of habituation, we asked participants to take the second section of the study at least one week apart after taking the first section. Each section includes Pre-then-post questionnaires. The pre-questionnaires includes questions about age, gender and nationality of the participants.

The post-questionnaires includes six questions, ranked from one to five. The goal of the post-questionnaires is to be able to compare two medians and find out the differences between them and if there is any differences, does physiological signals capable of showing this differences. Also the mean of answers can be considered as a score for the median that shows how strong the content is to arouse emotion.
Later on, the result can be compared to the result from physiological signals. Figure 4.3 shows post-questionnaires we used in our study.

4.3.2.3 Ethics Consideration

Because of the content of the stimulus, the participants have to be above 18 years old to be able to participate. The participation in the study is voluntary and at any stage of the study they can quit the study.

Although it is possible to provide stronger contents that might cause fear but for ethical purposes we decided to choose the content based on how fear is represented in entertainment industry.
Figure 4.3: Post-questionnaires for the study.
4.3.3 Data Collection

The bio-sensor that we are using for this study is E4 wristband by Empatica [21] (Figure 4.4). This device is capable of streaming Electrodermal activity at 4Hz, Blood volume pulse at 64Hz, XYZ acceleration at 32Hz and skin temperature at 4Hz.

![E4 wristband by Empatica.

Figure 4.4: E4 wristband by Empatica.]

One of the main advantage of using E4 compare to other devices is that the device is easy to use by users. This makes it possible for users to use the device with ease during daily activities.

This device is capable of measuring electrical conductance of the skin in the $[0.01, 100]\mu S$ (micro-Siemens) range with digital resolution of 1 digit per 900 $pS$ (pico-Siemens) [21]. As far as electrodes of the device are properly placed and the skin is healthy and hydrated, this resolution will be enough to sense EDA on young children and elderly peoples.

For receiving the data from Empatica to local machine, we used Bluegiga BLED112 Bluetooth low energy dongle [30]. Beside using bluetooth dongle, the local machine requires a
server to receive real time data from E4 to be able to process the data. We used TCP Server named Empatica BLE Server. The TCP Server is provided by Empatica and by the time of writing the thesis, it is on Beta state. Empatica TCP Server allows to receive real time data from multiple Empatica E4 Device on local machines.

To make sure the collected data is synchronized with the highlights of the stimulus, at each highlight we stored the clock time of the wristband and the machine that plays the stimulus. Later on, we use the clock time of the device to sync the data to the time of the highlights.

4.3.4 Data Processing

The analyzer receives synced raw data from the study, then different methods can be implemented for testing our hypothesis. Here we describe how the data being standardized and then cover our approach for the system to detect highlights of the video clip and finally our method for clustering the users is described.

4.3.4.1 standardizing the data

One problem with quantifying EDA signal is the existence of large variability because of individual differences. For example, the amplitude of $0.5 \, \mu S$ SCR might be high for one person and the same value might be the value of the baseline for another person.

To correct this inter-individual differences, one method is to standardize the data.

Standardization is the process of implementing and developing technical standards on data-
set to help maximize repeatability, or quality of the data. There is no universal approach on methods to standardize EDA signals.

For our system, we implement and compare the result from two commonly used approaches, *Range-Corrected Scores* [11] and *transforming raw data into Z-scores* [6].

**Range-Corrected Scores**: (Equation 4.1)

\[ X_n = \sum_{n=1}^{N} \frac{x_n - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \]  

(4.1)

Where \( X_n \) is N length standardized raw signal. Here N is the duration of stimulus. \( x_n \) is the value of raw signal at time n. \( x_{\text{min}} \) is the minimum value of raw data during baseline or relax period, and \( x_{\text{max}} \) is the value of the maximum aroused period.

**Z-Score**: (Equation 4.2)

\[ X_n = \sum_{n=1}^{N} \frac{x_n - \bar{\mu}}{\bar{\sigma}} \]  

(4.2)

Where \( X_n \) is N length standardized raw signal. Here N is the duration of stimulus. \( x_n \) is the value of raw signal at time n. \( \bar{\mu} \) is the mean of the raw signal. \( \bar{\sigma} \) is the standard deviation of the raw signal.

One problem with Range-Corrected Scores is that what counts as “minimum” might be the lowest value that the device can record and it might not necessarily represent users’ signal. Also, this statement is valid about the maximum value that, it is not clear if the value is artifact caused by device or it is the correct response caused by individual.

After standardizing raw SC data, We decompose it into phasic and tonic responses using the
Figure 4.5: Raw and standardized SC signal with phasic and tonic responses.

continuous decomposition analysis (CDA) method [4]. Figure 4.5 shows raw and standardized SC data with phasic and tonic responses.

4.3.4.2 Event Detection

In our study, we focused on EDA signal as the physiological signal for detecting highlight of the video clips that might arouse fear. Highlight is considered as events that are more common among majority of users. Although the system can show user-independent highlights but the output result shows low precision on detecting the highlight. The system provides different level of certainty that can detect overall highlights among all users with high precision and $F_1$ score.

Our approach for detecting the highlight is as follow:

The raw SC data needs to be standardized and decomposed into tonic and phasic responses.
It is recommended to standardized the data before decomposing the signal. The result of our study shows that, the standardized data has higher precision compare to non-standardized data. In our study we are using the continuous decomposition analysis (CDA) method to extract phasic and tonic responses [4].

We considered *Area Under the Curve* and *Sum of Squares*, with window size of three seconds [6] and moving window of 1 second, as possible features for detecting highlights. We explored raw, tonic and phasic responses from different users, and analyzed the output (Figure 4.5).

By comparing the result from different input signals, we decided to use *Area Under the Curve* for phasic responses as our main feature for detecting highlights. Figure 4.6 shows both Square Sum and Area Under Curve approach for standardized EDA, tonic and phasic signal.

![Figure 4.6: Square sum and area under curve comparison.](image)

After calculating this feature for all users, we detect and locate the peaks of the feature for all users. Figure 4.7 shows the detected peaks in red for 20 users.
The histogram of the calculated peaks can be used to locate at which time of the stimulus, higher phasic activities has been accrued (Figure 4.8).

Bins with higher values, show a larger group of people having higher phasic responses during bin’s duration. This information can be use to detect highlights in the stimulus. Figure 4.9 shows histogram with 50 bins including the participants inside each bin. The numbers on $X$ axis represent users and on the $Y$ axis is the location of the bin.

By adjusting the threshold for the number of users in the bins, we have a selection of bins that can be analyzed further more for detecting highlights. To analyze the bins, we calculate the Jaccard similarity coefficient between bins within the same level of users.

**Jaccard similarity coefficient** is a method for measuring the similarity of sample set. It
is defined as the size of the intersection over the size of the union of the set (Equation 4.3).

\[ J(A, B) = \frac{|A \cap B|}{|A \cup B|} \]  

The result from Jaccard dissimilarity is between zero and one, and the value closer to one means two sets share more common elements. By defining the threshold for Jaccard dissimilarity, we can select bins on the same level of the histogram that have some similarities to each other.

We define a graph to connect similar members on the same level together. First, we calculate Jaccard dissimilarity between each bin on the same level. By defining the threshold for Jaccard dissimilarity, we can exclude bins that are not similar to each other.

From this, for every similar bin, we find the intersections between two bins. By using stack
Figure 4.9: Users appearing inside the bins.

data structure, we push the intersections into the stack, then we pop one member and connect the member to the remaining items in the intersection and add the connected edge to the graph. We repeat the same process for all the bins in different levels.

The graph with the highest number of edges has higher probability to locate the time that might be considered as highlight. By defining a threshold for the number of edges in the graph, we can include bins that might include some highlights from the stimulus. Algorithm 1 shows the step taken to report the highlight.

Although it is possible to include all the level of bins from the histogram, but the better approach is to limit the number of levels to the numbers that makes sense based on the size of the collected data. In our study, we used the constant of five to include any bin size above or equal to five.

Figure 4.10 shows the result in Karate Graph, for the histogram of size 30 bins. In this example, bins with size bigger than four were considered as bins with potential highlights in them.
Algorithm 1: Reporting highlights.

**input**: Bins at the same level of the histogram, Jaccard Dissimilarity Threshold, Edge density Threshold

**output**: Possible location of the highlights

```
for bin ∈ InputBins do
    for bin ∈ InputBins do
        Calculate Jaccard Dissimilarity between bins
    end
end

graph ← new Graph()

while not at the end of bin ∈ InputBins do
    CurrentBin ← bin
    for bin ∈ InputBins do
        if jaccardDissimilarity(CurrentBin, bin) ≥ JaccardThreshold then
            intersection ← Intersection(CurrentBin, bin)
            while length(intersection) ≥ 1 do
                CurrentPerson ← intersection.pop()
                for members ∈ intersection do
                    graph.addEdge(CurrentPerson, members)
                end
            end
        end
    end

if graph.EdgeNumber ≥ EdgeDensityThreshold then
    Report location of the bins in the level.
```
Link below is a short video that describes the steps taken in the algorithm for detecting highlight:

https://youtu.be/BBX9fCH4G6I

![Figure 4.10: Output result.](image)

Table 4.3: Number of edges in each bin Level.

<table>
<thead>
<tr>
<th>Number of users in Bin</th>
<th>Number of Edges</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>11</td>
<td>48</td>
</tr>
<tr>
<td>12</td>
<td>58</td>
</tr>
<tr>
<td>13</td>
<td>60</td>
</tr>
<tr>
<td>14</td>
<td>28</td>
</tr>
<tr>
<td>15</td>
<td>103</td>
</tr>
<tr>
<td>16</td>
<td>69</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
</tr>
</tbody>
</table>
The density of the edges in each graph, shows the similarity between bins in that level. Bins with higher number of edges have higher probability to be considered as highlight of the stimulus, because higher number of similar group of people reacted to the events during that bin. Table 4.3 shows the number of edges in Figure 4.10

After defining the threshold for edge density, we can find the location of the bins with dense edges. The selected bins has high probability to include some highlights of the study. For our study we decided to select the three graphs with highest number of edges.

The three threshold that we use in this approach are as follow:

- Number of Bins in the histogram
- Jaccard dissimilarity threshold
- Edge density threshold for Karate graph

Different threshold levels can provides different level of certainty. For example, with low number of bins, low Jaccard value and low edge density threshold, the location of highlights may point to the whole duration of the stimulus. Figure 4.10, shows the output with total number of 30 bins, with 0.45 Jaccard dissimilarity threshold and edge density of 40.

4.3.4.3 Similarity Measures and Clustering

Clustering time series data differs from clustering of static features data mainly because of differences between methods for computing the similarities between two data objects. [31]
The key for clustering is to understand the unique characteristic of the data and design an appropriate dissimilarity measure method. We present a method to measure similarities between different users based on how common they react to the highlights of the stimulus.

First, we explain terminologies we used in the method:

**Pearson Correlation Coefficient:**

“Correlation” is a widely used concept to express the relationship between one quantity to another. There are different methods for calculating correlation such as *Pearson Correlation Coefficient, Spearman’s Correlation, Kendall’s Tau*, etc.

Pearson is the most widely used correlation method. It measures the linear correlation between two variables. It is represented by Greek letter $\rho$ and the formula for calculating the coefficient is (Equation 4.4):

$$
\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}
$$

(Equation 4.4)

where $\text{cov}(X,Y)$ is the covariance of two variable and for discrete variables can be calculated as follow (Equation 4.5):

$$
\text{cov}(X,Y) = \frac{1}{n} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})
$$

(Equation 4.5)

$\bar{x}$ and $\bar{y}$ are the means of the data and $n$ is total number of sample. $\sigma_X$ and $\sigma_Y$ are standard deviation of $X$ and $Y$. 
Dividing covariance by the product of standard deviations ensures that the correlation coefficient will always be in range of -1 to 1. This makes it possible to have a level of threshold for measuring the similarities between two variables.

**Inverse Frequency:**

In information retrieval the term inverse document frequency is used to measure commonness of specific term. The formula for calculating it is (Equation 4.6):

\[
idf_t = \log \frac{N}{n_t}
\]  

(4.6)

Where N is total number of items in the document and \( df_j \) is total number of the term \( t \) in the document. In our method we using the same concept to measure how common individuals react to highlights in stimulus.

To measure the similarity between individuals, first we calculate *Pearson Correlation Coefficient* for the duration of response window time at each highlight. Response window is the time window between stimulus-elicited and EDA response. It assumes to have a value between 1 to 4.2 seconds [6].

Then by defining a coefficient threshold, we can group different individual for each highlight. Different methods of standardizing raw data does not have affect on the value of Pearson Correlation Coefficient.

After calculating the coefficient, there will be another between group comparison to make
Figure 4.11: Grouped individuals in event.

Table 4.4: Groups in event.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Size</th>
<th>Members in Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>[1, 2, 4, 9, 11, 14, 16, 17, 18, 20]</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>[3, 6, 7, 8, 10, 12]</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>[5]</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>[13]</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>[15]</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>[19]</td>
</tr>
</tbody>
</table>

Sure each individual appears in one group only. For example, for a threshold of 0.85 for Pearson Correlation Coefficient, in one group we have a set of “user 1” and “user 2” with a Pearson Correlation Coefficient above 0.85 and in another group we have “user 3” and “user 2” with a Coefficient above 0.85, after in between comparison, the two grouped merges together.
Figure 4.11 shows the final result of the event and different groups of users in that event. By giving a score to each individual based on how common they appeared in groups of different events, we can cluster the result. The score is the value of inverse frequency of the individual.

The process of calculating inverse frequency is as follow:

In every event, each individual appears in one group, we count the size of the group the individual belongs to. The score of the user for that event will be $\log n$, where $n$ is the size of the group. If the size of the group is one, then the individual will take the score of zero.

For example, Table 4.4 shows the members of each group in “Person On Stair Scene Change” event. User with id number 2 will have a score of $\log 10 = 1$ and user with id number 3 will have a score of $\log 6 = 0.78$ for this event. If the value of the score is high for the individual then it means, that individual appeared more commonly in all events.

After finding the score for each participant, we can use different distance-based methods of clustering. For our study, we are using k-nearest neighbor to cluster participants.

Link below is a short video that describes the steps taken in this approach to find the similarity between signals based on the highlights:

https://youtu.be/ALkuzcKscMQ
A total of 20 users (female = 8, male = 12), with age between 18 to 40, participated in the study. During the study we measured EDA signal using E4 Empatica wristband and after study we asked the participants to fill the post-questionnaires (Figure 4.3).

For hypothesis testing, due to the small sample size and, therefore, uncertain underlying distribution of the data, we considered both parametric and nonparametric approaches for analyzing the data.

We used paired t-test for parametric and Wilcoxon test for nonparametric approach.

The result for both approaches shows that participants are more engaged and involved to the video game content compare to the movie (p-value < 0.05). Other than this, there are no significant differences between two contents.

Table 5.1 shows overall mean and standard deviation for different questions between movie
and video game stimulus. The questions were ranked between one to five, which value of 1 represent negative feedback to the question and value of five represent positive feedback.

Table 5.1: Mean and Standard Deviation between movie and video game.

<table>
<thead>
<tr>
<th></th>
<th>Movie Mean</th>
<th>Movie Std</th>
<th>Video Game Mean</th>
<th>Video Game Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>Being engaged and involved</td>
<td>3.40</td>
<td>0.99</td>
<td>4.00</td>
<td>0.79</td>
</tr>
<tr>
<td>Negative or Positive Feeling</td>
<td>2.80</td>
<td>0.69</td>
<td>2.75</td>
<td>1.07</td>
</tr>
<tr>
<td>Being Passive or Active</td>
<td>2.75</td>
<td>1.02</td>
<td>2.95</td>
<td>1.14</td>
</tr>
<tr>
<td>Being Relaxed or Tensed</td>
<td>3.50</td>
<td>0.94</td>
<td>3.10</td>
<td>0.23</td>
</tr>
<tr>
<td>Dislike or Like</td>
<td>3.15</td>
<td>0.18</td>
<td>3.15</td>
<td>0.26</td>
</tr>
<tr>
<td>Being in Control</td>
<td>3.20</td>
<td>0.28</td>
<td>3.20</td>
<td>0.26</td>
</tr>
</tbody>
</table>

Table 5.2 show the overall mean and standard deviation for all questions.

Table 5.2: Mean and Standard Deviation between movie and video game.

<table>
<thead>
<tr>
<th></th>
<th>Movie Mean</th>
<th>Movie Std</th>
<th>Video Game Mean</th>
<th>Video Game Std</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.13</td>
<td>0.30</td>
<td>3.19</td>
<td>0.42</td>
</tr>
</tbody>
</table>

We can conclude, overall, two stimulus are similar to each other and because of the unique characteristic of the content in the video game, for example using camera in first person perspective, the users might become more engaged to the content.

Beside this, the mean and standard deviation for “Being in Control” shows that the level of interaction for both stimulus are identical.

The overall mean (Table 5.2) for movie and video game can also be considered as the rank for how strong the content can be. Value closer to five means the content might have higher
emotion arousal on the users.

Overall, the result from self report shows, the content in both movie and video game are very similar together and there is not a significant differences between two contents.

For analyzing the EDA signal, we divide our results into to sections. First we cover the result from our approach for detecting the highlights of the stimulus, after that we discuss our result for clustering the participants

5.1 Highlight Detection

Sympathetic Nervous System (SNS) is part of the Autonomic Nervous System (ANS) that can often responsible for what is known as flight or flight response.

The phasic response of EDA signal is a parameter that can be used for detecting highlights of the stimulus. It has a response window between 1 to 4.2 seconds [4]. To effectively include the phasic response, we decided, to measure the area under the curve and square sum of the values with a window size of three seconds at every seconds of the standardized signal.
Figure 5.1: The area under the curve and Square sum for EDA, tonic and phasic response.
Figure 5.1 shows the result of square sum and area under the curve for three seconds for Z-score EDA signal, tonic and phasic responses for all data. As expected, the result from phasic response can be used for detecting highlights.

In some cases, the area under the curve showed more peaks with bigger amplitude compared to square sum. This difference between area under the curve and square sum can indicate that the area under the curve is more sensitive to the amount of changes that accrued in duration of three seconds.

From this result we decided to locate the peaks for area under the curve of phasic response (Figure 5.2).

![Graphs showing peak detection](image)

**Figure 5.2:** Area under the curve of phasic response with detected peaks.

Figure 5.3 shows the location of the peaks for all participants during stimulus.
We visualized this output using histogram with different number of bins. Figure 5.4 shows the histogram for all peaks with different number of bins.

The result from our approach for measuring the commonality between bins are displayed for different bins, using karate graph in Figure 5.5.
<table>
<thead>
<tr>
<th>Number of Bins</th>
<th>Movie</th>
<th>Video Game</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td><img src="image1.png" alt="Histogram" /></td>
<td><img src="image2.png" alt="Histogram" /></td>
</tr>
<tr>
<td>40</td>
<td><img src="image3.png" alt="Histogram" /></td>
<td><img src="image4.png" alt="Histogram" /></td>
</tr>
<tr>
<td>50</td>
<td><img src="image5.png" alt="Histogram" /></td>
<td><img src="image6.png" alt="Histogram" /></td>
</tr>
<tr>
<td>60</td>
<td><img src="image7.png" alt="Histogram" /></td>
<td><img src="image8.png" alt="Histogram" /></td>
</tr>
<tr>
<td>70</td>
<td><img src="image9.png" alt="Histogram" /></td>
<td><img src="image10.png" alt="Histogram" /></td>
</tr>
<tr>
<td>80</td>
<td><img src="image11.png" alt="Histogram" /></td>
<td><img src="image12.png" alt="Histogram" /></td>
</tr>
<tr>
<td>90</td>
<td><img src="image13.png" alt="Histogram" /></td>
<td><img src="image14.png" alt="Histogram" /></td>
</tr>
<tr>
<td>100</td>
<td><img src="image15.png" alt="Histogram" /></td>
<td><img src="image16.png" alt="Histogram" /></td>
</tr>
</tbody>
</table>

Figure 5.5: Histogram of peaks with different number of bins.
For each bin, we selected top three graphs with highest number of edges. From the selected graphs, we extracted the location of the bins in that graph. Finally, we compared the extracted location to our pre-defined highlights. We tried our approach with and without Jaccard threshold to be able to compare the result. The result shows that the Jaccard dissimilarity is not an important threshold to be used in this approach. What can be considered as important variable for detecting highlights is our approach for finding commonality between bins.

Table 5.3 and table 5.4 show the precision, recall and F1 score for each bins with and without including jaccard threshold.

Table 5.3: Highlight detection including all signals for movie.

<table>
<thead>
<tr>
<th>Number of bins</th>
<th>Jaccard Threshold = 0.0</th>
<th>Jaccard Threshold = 0.4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Precision Score</td>
<td>Recall Score</td>
</tr>
<tr>
<td>100</td>
<td>0.93</td>
<td>0.84</td>
</tr>
<tr>
<td>90</td>
<td>0.89</td>
<td>0.80</td>
</tr>
<tr>
<td>80</td>
<td>0.93</td>
<td>0.87</td>
</tr>
<tr>
<td>70</td>
<td>1.0</td>
<td>0.74</td>
</tr>
<tr>
<td>60</td>
<td>0.93</td>
<td>0.90</td>
</tr>
<tr>
<td>50</td>
<td>1.0</td>
<td>0.87</td>
</tr>
<tr>
<td>40</td>
<td>1.0</td>
<td>0.84</td>
</tr>
<tr>
<td>30</td>
<td>1.0</td>
<td>0.64</td>
</tr>
</tbody>
</table>
Table 5.4: Highlight detection including all signals for video game.

<table>
<thead>
<tr>
<th>Number of bins</th>
<th>Precision Score</th>
<th>Recall Score</th>
<th>F1 Score</th>
<th>Precision Score</th>
<th>Recall Score</th>
<th>F1 Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.96</td>
<td>0.67</td>
<td>0.79</td>
<td>0.92</td>
<td>0.7</td>
<td>0.8</td>
</tr>
<tr>
<td>90</td>
<td>0.93</td>
<td>0.75</td>
<td>0.83</td>
<td>0.9</td>
<td>0.78</td>
<td>0.84</td>
</tr>
<tr>
<td>80</td>
<td>0.96</td>
<td>0.78</td>
<td>0.86</td>
<td>0.96</td>
<td>0.81</td>
<td>0.88</td>
</tr>
<tr>
<td>70</td>
<td>0.92</td>
<td>0.83</td>
<td>0.87</td>
<td>0.96</td>
<td>0.78</td>
<td>0.86</td>
</tr>
<tr>
<td>60</td>
<td>0.91</td>
<td>0.86</td>
<td>0.89</td>
<td>0.89</td>
<td>0.95</td>
<td>0.92</td>
</tr>
<tr>
<td>50</td>
<td>0.96</td>
<td>0.75</td>
<td>0.85</td>
<td>0.97</td>
<td>1.0</td>
<td>0.98</td>
</tr>
<tr>
<td>40</td>
<td>0.96</td>
<td>0.67</td>
<td>0.79</td>
<td>1.0</td>
<td>0.62</td>
<td>0.76</td>
</tr>
<tr>
<td>30</td>
<td>1.0</td>
<td>0.73</td>
<td>0.84</td>
<td>1.0</td>
<td>0.7</td>
<td>0.84</td>
</tr>
</tbody>
</table>

To test our method, we ran the method for per-individual signals instead of including all the signals. Table 5.5 shows the average precision and F1 score among all participants.

Table 5.5: Highlight detection per individual.

<table>
<thead>
<tr>
<th>Movie</th>
<th>Video Game</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Precision</td>
<td>Average Recall</td>
</tr>
<tr>
<td>0.43</td>
<td>0.27</td>
</tr>
<tr>
<td>0.47</td>
<td>0.21</td>
</tr>
</tbody>
</table>

The result from our study shows that although it is not possible to detect the highlight for individual but by including all signals this method is capable of reporting the location of highlights.

The precision of this methods depends on the size of the bins, and number of edges in karate graph.
looking at the result, the two bins, number 60 for movie and number 50 for video game, shows Jaccard threshold caused improvement in F1 score. The selected threshold for Jaccard dissimilarity shows some improvement but overall, it is not significant.

By analyzing the time interval between each pre-defined highlights we have a mean of 9.85 with standard deviation of 4.51 for movie stimulus and mean of 8.65 with standard deviation of 5.70 for video game stimulus. By comparing the time intervals between each highlight and our finding, the result shows this approach has a potential to be able to detect highlights based on commonality with high F1 score.

This result show that, EDA signal by itself can be sufficient for detecting highlights of the stimulus without using other physiological signals. Beside this, phasic response plays an important role in detecting the highlights.

5.2 Similarity Measurement and Clustering

As mentioned above, our first approach for clustering the data is to group participants based on the value of the cross-correlation with response window of three at the location of each pre-defined highlights.

Because of the nature of EDA signals and individual differences in skin qualities, selecting a method for standardizing data is a challenging process.

One of the advantages of using cross-correlation for grouping different signals is the fact that
different methods of standardizing the dataset will not have effect on the final output.

Figure 5.6 shows the result of this process for one of the highlights of the study. Both Z-score and Range Corrected methods will give same result.

![StartRelax](image)

Figure 5.6: Grouped participants from different standardization methods.

This step can help us to receive same output with different standardizing approaches. From this result, we decided to use Z-score to standardize our data.
By comparing the members in each group from one signal variable to another, we were not able to find any relationship in-between groups of EDA, tonic and phasic responses.

We grouped participants that has a correlation above 0.8. Figure 5.7 shows the result of grouping participants based on EDA, tonic and phasic signal. The result shows that, the decomposed signals does not share the same grouping result with the EDA signal. From this, we decided to only use Z-score EDA signals for measuring similarities between different groups in the events.

Figure 5.7: Grouped participants.

After grouping participants for each event we calculated the Inverse Frequency for each
person and used the final score as a distance measure for clustering. We used a knn method \( k = 3 \), to cluster the users.

Table 5.6, shows participants in each cluster. The clusters with smaller center value, can represent users that share less commonality between other participants, vice versa, clusters with higher value can represent users with more commonality between them.

It is important to mention that, commonality does not mean that they might show higher arousal level to stimulus. This means users in the same cluster might appear in equal number of highlights during stimulus.

By looking at the Z-score EDA signals from the members in each group (Figure 5.8), it is possible to conclude that the overall shape of the EDA signal, for the majority of the participants in cluster number one and three are similar to each other. On the other hand, in cluster number two, the shape of the EDA signals does not follow specific pattern.

Further study with bigger sample size is required to provide a better conclusion from current approach but looking at the result of the clustered data in number three for video game, it might be possible to say that the common reaction to the content is a tonic responses without any strong phasic reaction. On the other hand, participants in cluster 2 for both movie and video game, does not follow any specific shape, and thus it can show that the role of individual in measuring emotion plays an important part.
Table 5.6: Clustered participants using $k = 3$.

<table>
<thead>
<tr>
<th>Cluster Number</th>
<th>Cluster Center</th>
<th>Participants</th>
<th>Cluster Center</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15.50</td>
<td>5, 13, 16, 19</td>
<td>19.67</td>
<td>1, 3, 4, 10, 18, 20</td>
</tr>
<tr>
<td>2</td>
<td>19.97</td>
<td>3, 4, 7, 11, 14, 15, 18, 20</td>
<td>23.15</td>
<td>2, 5, 6, 7, 8, 9, 13, 15, 16</td>
</tr>
<tr>
<td>3</td>
<td>23.61</td>
<td>1, 2, 6, 8, 9, 10, 12, 17</td>
<td>26.17</td>
<td>11, 12, 14, 17, 19</td>
</tr>
</tbody>
</table>

Figure 5.8: Clustered EDA signals.
5.3 Discussion

The overall result shows that, although the content of the “Halloween” is old but it can be used for further studies as a stimulus that can arouse fear.

There are no sign of large phasic responses in majority of signals which also shows that the overall rank for media (3.13 for movie and 3.19 for video game) might have positive correlation to the phasic response. From this result we might be able to predict that the EDA signal for majority of users might include only tonic responses without any strong phasic responses.

To explore in more detail to find out if the overall score for the content can be representative, a polynomial regression model with a degree of four were used based on the location of the local maxima from calculating the area under the curve for phasic responses (Figures 5.9 and 5.10).

The result from this regression model shows, There are more phasic responses in the movie stimulus compare to video game. One of the possibilities for having this differences between two content might be that the first section of the study for all participants was the movie stimulus. The content of the stimulus was unknown to the participants and this might cause more phasic responses. This response reduced during video game stimulus, because the participate might anticipated the content of the stimulus and thus there are less phasic responses to the video game stimulus.

Also, during study, we did not notify participants about the first two section of the stimulus
which is black screen for 10 seconds and relax part for 90 seconds. This caused participants to anticipate for any shocking moments that might accrue during study. This might caused the feeling of unknown that led to higher phasic responses. By exploring the regression model from movie, it is possible to show, after the ding sound caused by the flower, there is gradually increases in phasic responses.

As Feldman [19] mentioned in her study, the self report from participants tends to show lower arousal level compare to the analyzed physiological changes.
Figure 5.9: Regression model for movie.

Figure 5.10: Regression model for video game.
Affective computing is growing and the idea that one day application can make decisions for us without our cognitive front is not out of reach. Today, biometric sensors are becoming smaller with better accuracy that can easily be used in daily tasks. These advancement, require to explore new methods and approaches in affective computing.

The results from our study shows by using just a wearable device capable of storing only EDA signal, it is possible to detects the highlight of the stimulus. This opens a new door for creating a system capable of recommending the stimulus based on the current physiological changes of the users. For creating such system, we also proposed a clustering method based on how common users react to different highlights.

By implementing these two methods inside a real time system it might become possible to build affective computing software capable of making decision based on physiological changes.
It is also important to mention that measuring emotion is a complex process. Many variables needs to be considered to be able to measure emotion. Beside this the sample size of the study is another important factor in research topics such as this.

Although the result of the study shows that by only using EDA signal it is possible to detect the highlights of the stimulus, but it might be possible that by increasing the number of participants, the shape of the histogram tends to become symmetric. This will lead to case that the algorithm might include the whole duration of the stimulus as highlight, which is meaningless.

Beside this, because of the small sample size, it is not possible to provide an accurate conclusion based on the results we found from clustering. In both cases, study with bigger sample size is preferred.

Also, for detecting the peaks in area under the curve of phasic responses, a better method with higher accuracy is preferred. The current approach is capable of detecting the peaks but the output result shows this method was not capable of fully detect all the local maxima.

Beside this, in this approach, it might be possible to use the total number of peaks per individual as a variable for detecting the quality of the signal. Though in this study we did not extracted any noisy data from our dataset, but the ratio of total number of peaks over the duration of the stimulus, can be used to identify noisy data. The decomposed phasic response with too many peaks can be considered as a noisy data. More study needs to be done to test this claim.
In our approach for detecting the highlights, we manually pre-defined the highlights. While the result of our research includes all the events that we assumed might be considered as highlight, but another study can be conducted that the highlights are defined by users. This approach can provide a better insight to see if the system can detect highlights per individual.

In our research we mainly focused on fear. For future work, it would be useful to run the same study with other basic emotions to find the physiological differences between emotions.

The current approach for clustering system only uses correlation between signals. It is possible to add more features such as measuring Dynamic Time Warping, and also including decomposed signals into the clustering system as more feature and analyze the result.

Overall the result from this study, shows the potential of EDA signals that can be used, independently, in affective computing. Although we were able to gain a better understanding of physiological signals and their correlation to stimulus, but for a complex system such as human emotion with many variables, more research needs to be done to be able to create a system that understands human emotion.
Bibliography


