Doctoral Dissertation
Academic Year 2021

An Interdisciplinary Psychometric Method for Psychophysiological Tracking of Social Interactions and Social Dynamics in "Real-World" Environments

Keio University
Graduate School of Media Design

Dingding Zheng

A Doctoral Dissertation
submitted to Keio University Graduate School of Media Design
in partial fulfillment of the requirements for the degree of
Ph.D. in Media Design

Dingding Zheng

Dissertation Advisory Committee:

Professor Kai Kunze	(Principal Advisor)
Professor Hiroyuki Kishi	(Co-Advisor)
Professor Masa Inakage	(Co-Advisor)

Doctoral Dissertation Review Committee:

Professor Hiroyuki Kishi	(Chair)
Professor Masa Inakage	(Member)
Lecturer Jamie A. Ward	(Member, Goldsmiths, University of London)
Lecturer Yingjie Wang	(Member, NUFE)

Abstract of Doctoral Dissertation of Academic Year 2021

An Interdisciplinary Psychometric Method for Psychophysiological Tracking of Social Interactions and Social Dynamics in "Real-World" Environments

Category: Science / Engineering

Summary

This thesis aims to evaluate a proposed new approach of exploring social interactions and group dynamics in daily life scenarios through psychophysiological tracking and its combination with traditional psychological and sociological research methods. When it comes to objectivity and validity, the widely used evaluation psychometrics, e.g. self-report questionnaires, surveys, and trained observers rating videotaped subjects are not without their limitations. Psychophysiological sensing data feedback can offer information that cannot be extracted from traditional method such as questionnaires, interviews and observation. However, current psychophysiological studies regarding social science disciplines span only laboratory environments with big sensing equipment where the subjects are taken out of their natural social environment. Hypothetically, field research can cover this blind spot. However, current field research typically applies traditional methods as data sources. The option of applying psychophysiological sensing methods in field research is severely limited by a lack of suitable equipment comparable with the lab setups, thus limiting the information acquisition versus traditional methods. This work aims to introduce psychophysiological sensing methods into social interaction studies in "real-world" everyday environments using modern wearable devices that can track Heart Rate and ElectroDermal Activity (EDA). We conducted a couple of large-scale recordings with wearable psychophysiological sensing devices to test the proposed concept as well as the designed platform. Furthermore, we recorded 5 social events in "real-world" settings with 52 participants (143.4 participant-hour). The analysis of the psychophysiological data we gathered from participants showed group excitement levels, synchrony levels of the group and individuals, as well as the dynamics among the group. These results supported the theory that the proposed approach offers a new dimension for the evaluation of social behaviors and cognitive status in addition to traditional methodologies. In addition, by processing the recorded psychophysiological data as a collective group and as individuals within the social environments allowed us to explore the social reality in a mutli-layered perspective. Furthermore, the development of this methodology led to connecting more data patterns with behavior features, which allows us to predict human behaviors and relationships. Example applications include facilitating better social interactions and good qualities of interpersonal relationships

Keywords:

field research, physiological sensing, smart wearable sensing, human interactions, social dynamic, social behavior

Keio University Graduate School of Media Design

Dingding Zheng

Contents

List of Figures

List of Tables

Acknowledgements

Kai Kunze, Masa Inakage, Jamie A. Ward, Yingjie Wang, Hiroyuki Kishi for supporting and advising this research

Gheorghe Chernisov, for his extraordinary support and dearly care

Yan Wu for assisting my field research
Organizational support from BOBO Li from x++ and Gandalf Li. All the people participated in this work
The Gesit real project, hacking studio and KMD, especially those friends and colleagues whom helped with my research: Mio Sugimoto, Jiawen Han, Danny Hynds (thanks for proofreading), Yan He, Yulan Ju, Kanyu Chen

Wanhui Zeng for remote accompany while I'm writing my thesis
Han Wu and Feihong Zhou for years of prayers

Chapter 1
Introduction

In this study, I would like to introduce my research on a possible new way to track the social dynamics and interactions of interpersonal relations in daily scenarios. The communication barriers existing among different disciplines are no surprise [13, 14]. This is no exception to social science and technology fields, while social science fields have delays in technology adoptions [15–18]; likewise, the technology field also sometimes fails to apply the findings from social science fields [16, 19]. Although developments and applications for wearable sensing have been explored in the human computer interaction (HCI) field for quite a while, it does not share the same fashion of application in solid research of psychology and sociology. On one hand, it can be explained by the lack of collaborations/communications across fields, imposing serious barriers for the cross-pollination of ideas. On the other hand, there is a strong hesitance in the social sciences to adopt any new technology, resulting from the incredible complexity of the methodology development. Viability studies and proposed applications for wearable sensing platforms performed in the HCI field fall very short of the requirements that social sciences pose to any new methods. This work attempts to bridge the two fields and propose a methodology that employs technical advances from the HCI field and produces results suitable for interpretation by social scientists.

The main goal of this work is to look into whether tracking psychophysiological data during social interactions from field experiments with small wearable devices can gather meaningful information that reflects the social reality and group dynamics.

1.1. Motivation

In a BBC documentary titled Animal Odd Couples [20], it showed cases of animals being close friends with other animals that are not from their species. Wildlife biologist Liz Bonnin, the presenter of the documentary, introduced a story about a bear, a tiger, and a lion as they carried on their friendship into adulthood. The owner of the land said she tried to separate them when they were no longer cubs but it caused a tremendous negative impact on all of them. They became very depressed and refused to eat food until they were reunited with each other. The animal psychologist in the documentary explained that companionship was crucial for their well-being. In order to retain such companionship, the interactions between them was essential. In some of the animal cross-species friendships, the interaction and companionship even played a life or death role in their lives, especially for the orphaned animals. In psychology, this is described as the stress-buffering effect (see the detail at 2.1.5). As social beings, we humans share the same need for social interaction and companionship.

With the sudden strike of COVID-19, a huge amount of the population were asked to cut their usual social activities and keep social distance, if not separation, from their usual social connections. Ever since it began, we were forced to reevaluate the meaning of our social connections and adjust our interpersonal relationships with families, professional associations, and the communities we live in. Despite the slow efforts in mass-vaccination roll-outs, COVID-19 didn't stop, and follows us to this current moment. Sociologists and politicians suggest to be prepared to co-exist with COVID-19 and to accept this as the new normal for our society. What kind of new normal social life are we dealing with? Since early 2020, we have seen a domino effect on our daily life, such as the increase of mental health issues as well as suicide rates. How can we maintain suitable levels of well-being and prevent further mental health issues? How can we maintain healthy interpersonal relationships under such circumstances?

The problems we face now were well known and predicted by academics much earlier. Yet, in daily life we are too distracted with other problems and not enough attention is spared to look at how crucial relationships in social life are [21]– since it can impact the mental, physical, purposeful, and almost every other aspect of our life [22–26]. The influence of the pandemic changed the old dynamic and the environment where we maintain or develop interpersonal relationships, compared to what we were used to. These changes revealed the existing physical and mental health issues of our well-being, raised the importance of our interpersonal relationships to higher priority, and pushed us to face problems that used to be deliberately well hidden. As noted by Gardner and his colleagues, successful interpersonal relationships and social interactions are crucial factors of the quality of human life.

1.2. Research Questions

The central questions of this thesis are as follows:

1. *Whether the proposed approach– applying psychophysiological sensing into field research in studies, such as social interactions and social dynamics, add a new dimension or offer new information to social interaction studies.*

2. *What are the requirements of the psychophysiological sensing system in order to be able to apply it in field research?*

3. *What are the requirements of the field setting for this work to record social interactions of people in "real-world" environments?*

4. *Can we get valid data through the psychophysiological sensing platform?*

5. *What kind of information can we extract from it that may not be available in applying the traditional method alone?*

These questions are practically and theoretically important for understanding whether and how well the proposed method will work in social science research, especially when it relates to interpersonal relations. In this thesis, we will cover these questions in the following chapters.

1.3. Thesis Outline

The outline of this thesis is as follows:

- Chapter 1 briefly introduces this thesis to the reader.
- Chapter 2 introduces the literature overview that supports the motivation, the problem statement, and the ground truth of the chosen approach for this work. It gives a broad overview of the psychological background of this work and the traditional research design in psychology as well as the relevant sensing approach.
- Chapter 3 describes the conceptual development of this work as well as specifications and requirements for testing the proposed approach by discussing the reasons for the shortcomings of current psychometrics and possible solutions with some concept testing studies. The problem discussions are based on chapter2.
- Chapter 4 describes the development of the custom platform of the proposed approach with some studies that motivate certain design decisions.
- Chapter 5 Presents the method of 5 field studies that applied the proposed approach and recorded the datasets.
- Chapter 6 demonstrates how the data can be meaningfully interpreted by adding psychophysiological sensing data together with traditional methods.
- Finally, Chapter 7 concludes this work. It summarizes everything and discusses the future prospects and potential applications of this work.

1.4. Contribution

This work is an interdisciplinary approach of studying social interaction and social dynamics in "real-world" scenarios. This work combines metrics from the disciplines of psychophysiology, field research, and computational psychometrics together with wearable computing and applies them in social interaction studies. In the social interaction field research, we recorded 52 participants in 5 social events hosted in 3 different cities and gathered useful data that delivered meaningful results. We recorded 143.4 participant hours including 1.1 GB psychophysiological data and 26.2 GB video recording. In each event, we recorded around

10- 20 people simultaneously. To our knowledge, the work covered in this thesis had not yet been done before in social interaction studies in psychological field research in this fashion.

This thesis:

1. presents an evaluation of the methodology of applying psychophysiological tracking to field studies of social interaction together with traditional psychometrics and found it to be a worthwhile approach.

2. Has developed a platform for such field studies and has developed combinations of different metrics to process the collected psychophysiological data for interpretation in the context of cognitive processes and social interactions.

3. Has offered all the recorded psychophysiological datasets for public and related communities with the interest of future analysis.

Chapter 2
Prior Art

In this chapter we will cover some explanation of the psychological and physiological concept that we used in this work. We also discuss about the related fields we built our method on.

2.1. Psychological Background and Concepts

In this section we will cover some definitions and concepts in the field of psychology and sociology.

2.1.1 Social Interaction and Social Dynamics

Social dynamics may have a different scale of definitions depending on the scale of the research topics. Auguste Comte conceptualized social dynamics differently from social static which represents the present structure of the society as specific societies and social systems in the process of historical change [27,28]. The American Psychological Association (APA) dictionary of psychology also explained social dynamics as "the forces or processes of change at work in any social group [28]. It also forwarded to the term of group dynamics, introduced by social psychologist Kurt Lewin [29] "the processes, operations, and changes that occur within social groups, which affect patterns of affiliation, communication, conflict, conformity, decision making, influence, leadership, norm formation, and power" [28]. In this work, based on the above definitions we consider social dynamics to be the collection of the behaviors and interactions between the individuals in all configurations: one to one, one to many and many to one. E.g. Two people talking (one to one); A person giving a speech (one to many); a group interviewing one person (many to one). In this work the social dynamics we are interested in are in the scale of the behavior of or within groups that results from the interactions of individual group members as well the relationships between individual actors and group level behaviors [30].

Sorokin defined social interaction as "any event by which one party tangibly influences the overt actions or the state of mind of the other" [31]. The APA dictionary of psychology explains social interaction as "any process that involves reciprocal stimulation or response between two or more individuals " [28]. A sociology textbook describes social interaction as the process of reciprocal influence exercised by individuals over one another during social encounters [32]. In this work, we tracked social interaction in the context of how individuals choose to act and react to other members within a social group.

Neuroscientists believe human functions, as well as humans brains and minds, are shaped in continuous interaction with other people [5]. Psychological findings do agree with this. In the case study of an isolated girl who was almost entirely deprived of social interaction for first 13 years of her life, there were severe delays in motor abilities and cognitive abilities development [33, 34]. Even after she got exposed with language input, like other similar cases, she failed to use language to communicate fluently [12, 35].

Moreover, studies have found social interaction is critical for psychological well-being, stress regulation, protection from disability and overall life satisfaction [5, 36]. Among all stages of adulthood, social interaction contributes a positive impact on psychological well-being [37].

2.1.2 Mental Health and Unawareness

The World Health Organization (WHO) estimated in 2016 the almost 800,000 people died from suicide (see as figure 2.1) and over 700,000 (with uncertain interval 516 000 to 966 000) in 2019 [38] . The WHO further reports more than 90 % of suicide cases are associated with mental health disorders [1] .

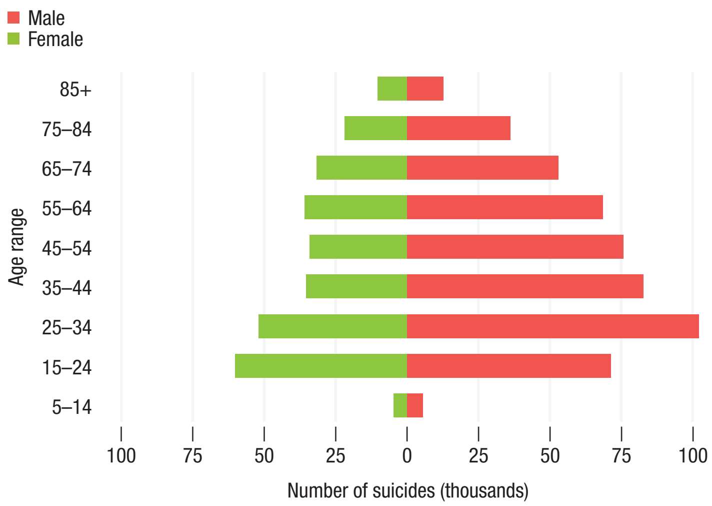

Figure 2.1 Global suicide deaths by age and sex, 2016 [1]

Andrade and his colleagues find the major barrier for people to seek mental health treatment is failing to recognize their need for treatment [21]. Their study examined 63,678 samples over 24 countries. They found the main barrier to initiation and continuation of mental health treatment among individuals with common mental disorders was perceiving a need for help. 63.8 % people wanted to handle it on their own; 24.4% of them thought they were not that severe; 16 % thought it would get better automatically. In other words, in many cases people who have long term negative emotions can be unaware of their mental health conditions.

2.1.3 Interpersonal Relations and Happiness

It is not a surprise that happiness is linked to a lot of important outcomes. Studies found happier individuals tend to have better a immune system, longer lifespans, better coping abilities, superior work outcomes, richer interpersonal relationships, and stronger social support [22–25]. Happiness has consistently been a subject people discuss and try to reveal its truth in the human history. In this section, we will discuss whether or how interpersonal relationships could impact one's happiness levels.

In order to study happiness in a quantitative empirical method, scholars tried to define happiness in a measurable way. In the research field, happiness has a synonymous term “subjective well-being” which is more clearly defined and easier for measurements. Lyubomirsky, Sheldon and Schkade [39] defined happiness “in terms of frequent positive affect, high life satisfaction, and infrequent negative affect” which are also the three primary components of the subjective well-being. Based on the definition of subjective well-being, a number of researches have designed self-report scales to measure “happiness”.

The quality of social relationships has been demonstrated as one of the most consistent predictors of happiness [26]. In a one semester long study, Diener and Seligman found comparing to less happy college students, the top 10% very happy students had stronger romantic and other social relationships. The results suggested good-quality social relationships may be a necessary condition for high happiness levels [40].

Longitudinal datasets of the over 80 years on-going Grant Study [41] and Glueck study [42] run by Harvard Medical School suggest that it is the relationships we have with other people that contribute the most to our well-being levels throughout our lifetime. As human beings, we need successful interaction with each other and with ourselves, as it is crucial for us and our relationships.

2.1.4 Circumplex Model: Valence and Arousal

According to James Russell Circumplex Model [43], emotions are distributed in a two-dimensional circular space, containing arousal and valence dimensions [43].In this model, emotional states can be represented at any level of valence and arousal, or at a neutral level of one or both of these factors. In order to enhance the happiness of people, there are two major directions: minimize the overwhelming negative emotions and increase the positive emotions.

2.1.5 Stress-buffering Effects

The stress-buffering hypothesis was firstly proposed in 1976. John Cassock [44] and Sidney Cobb [45] argued that strong social ties could protect individuals from the potential pathogenic effects of stressful events. Recent psychophysiological studies also described this phenomenon as a regulatory effect on individuals’ affective and physiological responses to distress. Especially for participants with secure attachment styles, the findings suggested that the presence of attachment figures tend to attenuate stress-induced ANS activity [46]. It is not to our surprise that supportive social interaction can relieve our psychological stress. Nevertheless, Uchino and his colleague revealed that over the life span, the stress-buffering function is a key mechanism through which social support promotes mental and physical health chronically [47].

Interpersonal relationship and Attachment

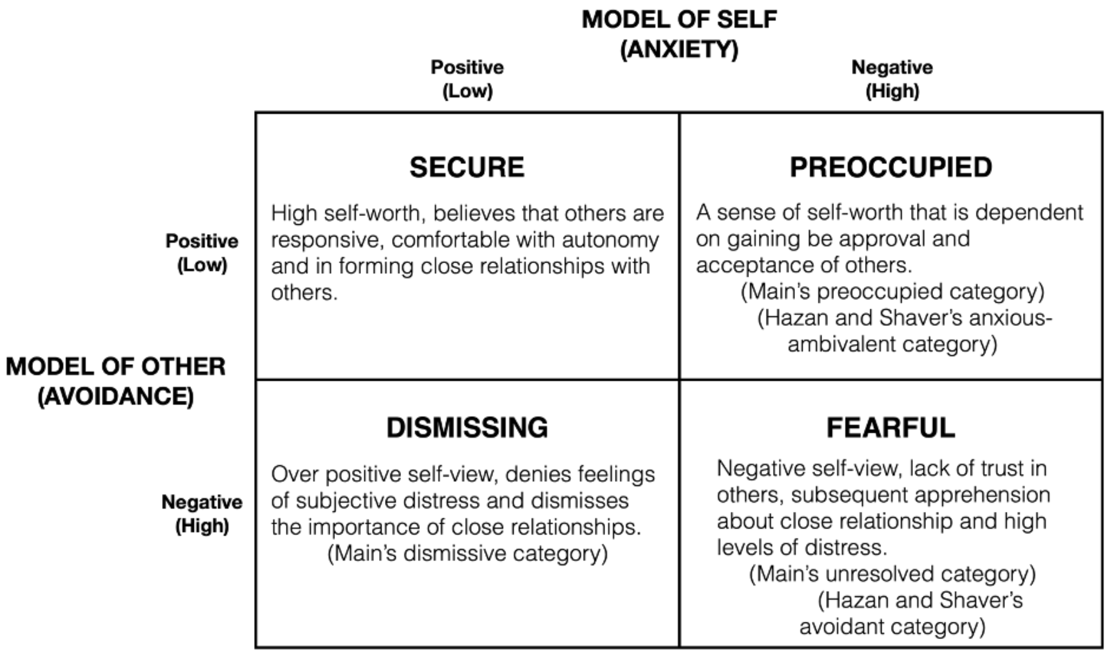

Figure 2.2 Bartholomew's Model of Four Categories of Adult Attachment [2]

Attachment styles were firstly introduced in infancy studies suggesting that there are 3 different attachment styles between a main caregiver and the baby [48]. Later on the inquiry was expanded to caregiver-child attachments in childhood, adolescence and adulthood. The attachment style with caregivers such as parents were found consistent across this time [49].

Developmental psychologists also found the same pattern in adulthood with interpersonal relationships towards romantic partners, family members, friends, and even co-workers. The researchers define adult attachment as the adult's current state of mind with respect to attachment based on early care-giving experiences [50]. Among all interpersonal relationships in adulthood attachment, romantic partners usually take over the place of caregiver and become the major attachment figures.

Based on the dependency levels and the avoidance of intimacy levels, researchers try to group up the attachment pattern in adulthood [2, 51, 52]. For instance, a low dependency pattern means a positive self-regard is established internally and does not require external validation; while a high dependency pattern suggests a positive self-regard can only be maintained by others' ongoing acceptance. The evaluation of avoidance of intimacy levels reflects the degree to which people avoid close contact with others as a result of their expectations of aversive consequences. Different attachment models are used and discussed. Most of them are based on two major patterns; secure attachment type and insecure attachment type. Since the insecure attachments are more diverse in patterns, there are more particular groups discussed in attachment studies. Here is one of the widely used model of adult attachment:

- Secure: Comfortable with intimacy and autonomy
- Preoccupied: Preoccupied with relationships
- Dismissing: Dismissing of intimacy, counter-dependent
- Fearful: Fearful of intimacy, socially avoidant

Most studies found significance for attachment figures such as romantic partners, while close friends and strangers appeared to be the same [47,53–55]. Several studies have found that the presence of supportive strangers is just as effective at buffering physiological stress reactivity as the presence of good (or best) friends, even when friends are subjectively judged as more supportive [53–55].

The loss of an attachment figure brings about distinct psychological deficits that cannot be attributed solely to the loss of partner support and that cannot be remedied by drawing on ancillary sources of support. Attachment figures are important and often preeminent support providers. Their influence on psychosocial and physical functioning should be considered independently from that of other support providers.

Among adults, men who performed a stressful task and who received assistance from their romantic partner showed less cortisol reactivity than those who received assistance from a stranger or who did not receive any assistance, although women did not show this pattern [56]. In other words, support from a romantic partner is more critical to a man compared to other relationships while a woman receive support from more diverse kinds of relationships.

2.2. Traditional Research Methods in Psychology

Taking psychology research as an example of how social science traditionally designs a research, there are two main types of research methods: experimental research and non-experimental research [57]. Experimental research focuses on causation, while non-experimental research very often uses methods such as naturalistic observation, case study, and correlational design.

- An experimental design involving random assignment of participants to conditions, manipulation of an independent variable (the presumed cause), measurement of dependent variable (presumed effect) and focus on draw cause-and-effect conclusions [12, 57].

- Naturalistic observation often uses a video camera or tape, sometimes even pens and papers to record and participants' behavior in real-world settings without trying to manipulate the situation [58].

- Case study design examines one person or a small number of people in depth, often over an extended time period [59].

- Correlational design examines the extent to which two variables are associated. Correlational designs can be extremely useful for determining whether two (or more) variables are related. As a result, they can allow us to predict behavior [12].

Table 2.1 Advantages and Disadvantages of Research Designs [12].

Methods	**Advantages**	**Disadvantages**
Naturalistic Observation	• High in external validity	• Low in internal validity; • Doesn't allow us to infer causation
Case Studies	• Can provide existence proofs; • Allow us to study rare or unusual phenomena; • Can offer insights for later systematic testing	• Low in internal validity; • Doesn't allow us to infer causation
Correlational Designs	• Can help us to predict behavior phenomena	• Don't allow us to infer causation
Experimental Designs	• Allow us to infer causation • High in internal validity	• Can sometimes be low in external validity

Field research, one of the main research settings we will be discussing in next session, is very often correlational studies. Correlational studies allow us to establish the relationships among two or more measures but do not allow to draw causal conclusions. Although correlation doesn't necessarily mean causation, a correlation sometimes reflects a causal relationship.

2.2.1 Laboratory vs. Field Research

There are two main types research settings: laboratory environments or real life environments. A laboratory study is a study that is conducted in a laboratory environment. In contrast, a field study is a study that is conducted in a real-world environment [57]. Since researchers have more control in laboratory environments, usually it is categorised as having high internal validity whereas field studies are categorised as having high external validity. Internal validity means to which degree a causal relationship between variables can be confidently inferred. External validity means to which extent the findings can be generalized in the real-world environment. In typical cases, making choices between laboratory environments and real-world environments are considered as a trade-off between internal and external validity [57].

Laboratory

Well designed laboratory studies have the advantage of highly controlled conditions where accurate measurements are possible. However, laboratory studies are typically low in external validity, while field studies are typically high in external validity.

Field

A field experiment is still an experimental design. Although it happens in the real life environment, the experimenter still manipulates the independent variables and tries to conclude a causal relationship between variables. Since the experimenter cannot really control extraneous variables, there are more alternative explanations besides the intended causal relationships between independent and dependent variables which lowers the validity of the conclusion.

A field study, also called as natural experiment, is a non-experimental design research in a natural environment. The experimenter has no control over the independent variables as they occur naturally in real life. This method is quite often used in longitude studies to explore what correlates to aspects such as well-beings, relationships of human in long terms.

2.2.2 Measuring Methods

Psychological measurement is often referred to as psychometrics [57]. Traditional data gathering methods of psychology studies include questionnaires, interviews, and observation methods such as rating videotapes by trained viewers [60]. Most of them quite often use self-report measures from the participants themselves or reports from people around them who have access to observe the participants well enough (frequently and in long term).

Questionnaires

Questionnaires are sets of questions, usually printed on paper or done on an electronic device. The subjects are required to answer them selecting one of the provided answers or put their answer on a scale. Questionnaires are valid only if the subject fully understands the questions and is self-aware enough to give an honest answer. Yet, the investigators usually have no control nor a way to know if the collected data fits the requirement or not.

Interviews

Interviews can be classified as either qualitative or quantitative. Qualitative interviews resemble a normal conversation and mostly utilize open-ended questions. The interview itself is either recorded as audio/video for later analysis or the investigator has to take notes or rely on their own memory. This kind of interview requires the investigator to understand the interviewee's behavior and utterances, and interpret them correctly. Quantitative interviews are basically surveys where the investigator puts down the answers based on own understanding and interpretation of the subject's answers. Thus, it raises the difficulties in replicating the experiments and results for peer scholars.

Observation

The observations face similar complications as the qualitative interviews from the data gathering view. This approach is often used in laboratory experiments or in field studies. The investigator is required to observe the behaviors and later analyze it based on the notes, their own memory and/or recordings. This method requires investigators to be well trained and have a certain level of expertise in detecting and judging certain behaviors with standard evaluation methods. It shares the same issues of interview method for replicating the same results.

2.2.3 Review of the Limitations

Self-report measures and surveys typically assume that participants are honest in their responses. Studies found that some respondents distort their answers to questions often in a way that paints them in positive light tendencies [61, 62].

It has a clear weakness due to the subjective nature of the data. First, they usually assume that respondents possess enough insight into themselves to report on feedback accurately [63, 64]. This judgement can be biased by individual differences such as personalities [65]. Studies suggested people with high levels of narcissistic personality traits view themselves more positively than others do [66].

In order to gather meaningful data, it is important to use questionnaires of high validity, train observers and interviewers to be less biased. These traditional data gathering methods have limitations as noted below:

- Very time-consuming to design and test a valid questionnaire as well as train observers and interviewers.
- Very expensive for the same reasons as above.
- Data is limited by subjective judgements of participants or investigators.

Some of these can be remedied by using psychophysiological sensing in addition to the traditional data gathering methods. Sensing devices collect straightforward data on the physiological responses that don't rely on the paricipants' self-awareness and judgements.

2.3. Psychophysiology

Psychophysiology is a field focuses on the interrelationships between mind and body. Cacioppo et al. define it as "scientific study of social, psychological, and behavioral phenomena as related to and revealed through physiological principles and events in functional organisms" [67]. An important characteristic of psychophysiology is it's top-down approach to the study of organismic-environmental transactions, complimentary to the bottom-up approach of psychobiology. Although psychophysiology as a field of study was formalized in the mid-20th century, early conceptions of human psychological and cognitive activities as embodied phenomena can be traced back to the works of Claudius Galenus from second century A.D. Galenus believed that mind and body were one and that this would be empirically demonstrated through science [68]. Galenus' understanding of mind and body defined the scientific views on psychophysiology for the next 15 centuries, until the scientific thought and measurement tools were advanced enough to start questioning Galen and give rise to the modern understanding of the connections between our mind and body.

Although the field of psychophysiology utilises a vast lineup of measurement tools, this work only focuses on the ones suitable for daily use in small form-factor. Thus, the measurement tools have to be reasonably small, easy to use and be safe even if misused. With this in mind we will not discus complex lab equipment used for e.g. neuroimaging (MRI, PET) as well as any other tools requiring highly trained medical personnel to be used. In the following pages we will discus certain physiological functions (e.g., sweat gland activities, peripheral blood flow) that covary with psychological processes (e.g., emotion, cognition, perception), focusing on the functional description of these phenomena and how monitoring psychophysiological responses in real time can help us to understand our thinking and feeling at the moment.

2.3.1 ElectroDermal Activity (EDA)

ElectroDermal Activity (EDA), also referred to as Skin Conductance (SC) or galvanic skin response (GSR), refers to the change of the electrical conductance properties of the skin in response to the change of the sweat secretion rates by sweat glands [10, 69]. EDA measurements mostly concentrate on two parameters: Skin Conductance Response (SCR) - quick changes (on the scale of seconds) in response to emotional or stress stimuli; and Skin Conductance Level - slow changes (within minutes and hours) commonly associated with the general condition of the subject.

A very common knowledge application scenario, in addition to the massive body of research in laboratory related to skin conductance and emotions, is that the modern polygraph systems (also known as lie detectors) rely on it to assess the unconscious emotional response and its effect on the physiological readings, as a reaction to the investigator's questions.

ElectroDermal Activity (EDA) tracking has a very long history in psychological research [70]. One of the first mentions of EDA usage for psychophysiological research was Carl Jung's book "Studies in Word Association" published in 1906 [71]. Nowadays it remains one of the very widespread tools in psychology and psychotherapy for measurement of autonomic nervous system responses [70, 72]. In recent decades, Skin Conductance (SC) is one of the most sensitive markers and frequently used to assess emotional arousal, as the skin conductance response activity increases as the emotional arousal grows. [10, 73–75].

2.3.2 Heart Rate Variability (HRV)

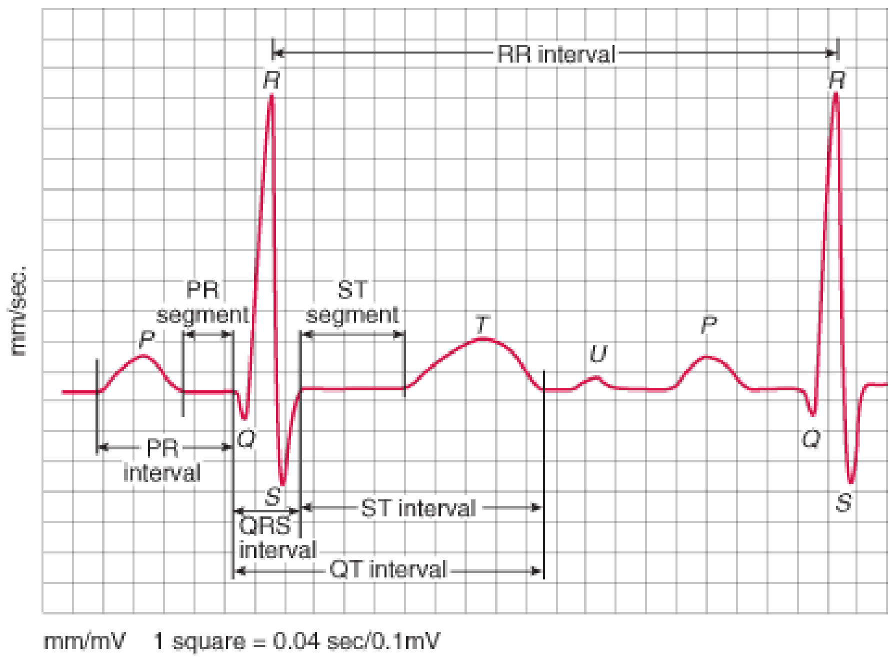

Figure 2.3 The RR interval is the time between two subsequent R-peaks [3].

HRV describes the changes in time intervals between each consecutive pair of heartbeats [76]. HRV is based on the analysis of the patterns in the Inter-Beat Intervals (IBI), also referred to as the RR-interval. The RR-interval is the time gap between two consecutive R-peaks in terms of a traditional cardiogram (see as Figure 2.3).

The earliest references to Heart rate variability (HRV) go back to 1733, when the Rev. Stephen Hales noted that respiration effects the heart contractions and blood pulsation. Carl Ludwig was the first to ever record and define the respiratory sinus arrhythmia (RSA) in 1847. RSA characterizes the increase in heart rate during inhalation and its decrease during exhalation [77].

Multiple studies address Heart Rate Variability (HRV) and emotions. A study by Choi and his colleagues [78], used HRV to evaluate the response to emotions induced using Affective Picture System (IAPS). They found out significant positive correlation of R-R interval (RRI) with valence and significant negative correlation with dominance, associated with the "unhappy" emotion. But only when the arousal value exceeded a certain value. It is suggested that it is possible to use an HRV-based evaluation for high arousal emotions. To clarify, R-R intervals are the time periods in milliseconds between two consecutive normal heart beats. By monitoring the heart rate using ECG or PPG, we can get a series of R-R intervals. The variance of the duration of those intervals is referred to as Heart Rate Variability (HRV). Although even without the variability analysis we can extract average it to get heart rate, the number of beats per minute.

HRV is often used to evaluate the activity of the autonomic nervous system, namely two of it branches, the sympathetic (SNS) and parasympathetic nervous systems (PSNS). The relation of SNS/PSNS, heart activity and underlying neural circuitry modulating these processes is described by the neurovisceral integration model [79]. Figure 2.4 depicts the detailed representation of the neurovisceral integration. Simply put, it demonstrates how the neural structures of the prefrontal cortex regulate the activity in limbic structures which modulate the balance between PSNS and SNS by inhibiting the PSNS and activating SNS neural circuitry. Since the prefrontal cortex activity is tightly linked with cognitive and psychological processes, it is possible to see the reflections of cognitive and emotional states of a subject in its HRV.

As described in table 2.2 SD of all N-N interval (SDNN) and the root of the mean square difference between adjacent N-N intervals (RMSSD) are two HRV time-domain measures. When the value of these HRV features go up it indicates the participant is getting more relaxed. In contrast, when they go down, the participant is getting less relaxed and more tense, as it signifies the activation of

Table 2.2 A Brief Review of Heart Rate Variability Measurements

Variable	Units	Definition
TIME DOMAIN MEASURES		
HRV triangular index		Number of normal R–R intervals divided by the height of the histogram of all the normal R–R intervals measured on discrete scale with bins of 1/128s (7.8125ms)
Statistical:		
SDNN	ms	**The SD of all normal R–R intervals**
SDSD		The related SD of successive R-R interval differences, only represents short-term variability
RMSSD	ms	**The root of the mean squared differences between adjacent normal R–R intervals**
NN50		The number of pairs of adjacent normal-to- normal R–R intervals that differ by more than 50ms calculated over over short period time (2 min)
pNN50	%	NN50 divided by the total number of normal R–R intervals×100
SD1	ms	The SD of Poincare plot perpendicular to the line-of-identity
SD2	ms	The SD of Poincare plot along the line-of-identity
FREQUENCY DOMAIN MEASURES		
Total	ms2	Area under the entire power spectral curve (usually 0.40), variance of all normal R–R intervals
LF	ms2	Low frequency power (0.04–0.15Hz), usually recorded over 2 mins
HF	ms2	High Frequency power (usually 0.15–0.40Hz), usually recorded over 1 min
LF/HF		Ratio of the low-to high frequency power

sympathetic nervous system moderating the fight-or-flight response [80]. Below are several standard HRV measures with their measurement units and respective descriptions. Standard metrics rely only on normal to normal beat intervals, meaning that abnormal heart beats caused by arrhythmia or other diseases or artifacts in the recordings are not counted.

The following are features and concepts that are related to this work but not directly covered in the table.

- **bpm**, beats per minute (bpm) is the number of heartbeats detected during one minute.

- **RSA** (respiratory sinus arrhythmia) is the naturally occurring variation in HR that occurs during the breathing cycle. During inspiration (breathing air in) the R-R interval on an ECG is shortened and during expiration (breathing air out) it is prolonged [81]. In other words, heartbeats synchronize with respiratory rhythm or breathing rate. Respiratory sinus arrhythmia (RSA) refers to the periodic fluctuations in heart rate [82]. RSA is directly proportional to HRV [83].

- **SD1/SD2**, Ratio of SD1-to-SD2, is a HRV non-linear measures. The ratio of SD1/SD2, which measures the unpredictability of the RR time series, is used to measure autonomic balance when the monitoring period is sufficiently long and there is sympathetic activation. SD1/SD2 is correlated with the LF/HF ratio [84, 85].

2.3.3 Social Interaction and Heart Activities

Shahrestani and her colleagues did a meta-analysis of 787 subjects from fourteen studies where the HRV data was collected during the participants in social interaction [36]. Ten of these studies used RSA as measurement, two used HF HRV, one used RMSSD, and one used SDNN.

According to Beauchaine's Integrated Model of Autonomic Functioning, low RSA characterizes negative emotional states such as anxiety, anger and depression [86]. Porges proposed the polyvagal theory, which concluded shifting emotional state covary RSA [87]. More specifically, a change to a more positive emotion associates with an increase in RSA ; whereas, a shift to a more negative emotion elicited by stressors associates with a reduction in RSA [82, 87]. In addition, Butler and her colleagues also addressed deceleration of heart rate along with an increase in RSA associates with a relaxation state in a safe environment; on the contrary, a decrease in RSA happens in an environment with the fight-or-Flight response [82].

Ditzena and her colleagues found during the Trier Social Stress Test (TSST) that was designed to induce stress of the participants [88], women who received a warm touch from a supportive partner associated with lower heart rate compared to a control group with no partner interaction [89]. Porges proposed the Polyvagal perspectives which addressed the relationship between HRV and emotion expression as well as social engagement [80]. According to the polyvagal perspectives when individuals are in a relaxed or resting state, HRV is generally increased, reflecting greater parasympathetic activity [80]. Shahrestani and her colleagues concluded negative social interactions lead to a decrease in HRV similar to the TSST [88].

2.4. Psychophysiological Sensing in Psychological Studies

Psychophysiological sensing has a long history in psychology studies. In this section, we give an example of how psychophysiological sensing is applied in psychology studies such as interpersonal relationship in Lab environment.

One big advantage of gathering data from physiological responses is that people can't consciously control many of the physiological functions. Therefore we can get objective data from sensing equipment. Another strength discussed in section 2.3 is that physiological functions connect with psychological process, allowing us to gather psychological information from collected data such as emotional arousal if applying EDA or HRV.

Two specific biological systems that have been found to be associated with attachment-related phenomena in prior social psychophysiological research: the autonomic nervous system (ANS) and the hypothalamic-pituitary-adrenal (HPA) axis of the endocrine system. A numbers of studies have used ANS and HPA for psychophysiology studies in the lab. Neither of them can not be consciously controlled. They have been widely tested in the theory of stress-buffering effect in which the investigator creates a stress event for the participant to see whether his or her stress levels will decrease or not with the attendance of an attachment figure, close friends, and strangers.

By applying psychophysiological sensing for data collection, studies extract information leading to meaningful findings as mentioned further on. Attachment studies that use psychophysiological sensing are good examples of how such methods can contribute to research with valuable data that other methods may not be able to.

By tracking psychophysiological responses during stress-buffering experiments, a number of studies have found that the presence of supportive strangers is just as effective at buffering physiological stress reactivity as the presence of good (or best) friends, even when friends are subjectively judged as more supportive [46]. The loss of an attachment figure brings about distinct psychological deficits that cannot be attributed solely to the loss of partner support and that cannot be remedied by drawing on ancillary sources of support.

However, these kinds of studies have to limit the interactions between the participants; otherwise the response gathered will be far from how people usually interact in everyday life scenarios. If we want to understand how people respond to social interactions in the "real-world", it is necessary to move the research setting to the field.

2.5. Field Research of Psychology Studies

Field research is the terminology often used in psychology research [57] and is popularly used as a "study in-the-wild" in the human computer interaction field (HCI) [90]. Field research are studies that are conducted in everyday life environments (i.e. real-world settings). This is also one of the examples of the challenges that interdisciplinary collaborations face– the experts speak the language of their own discipline, but can hardly navigate in other fields. [13, 14].

Field research is important in social science not only because the setting is in the natural environment, but also in some research topics such as relationships done in a lab setting, could lead to ethical issues. For instance, the investigators very often deliberately create a stressful or intense situation for participants attending the research in order to test their responses such as stress buffering effects. If investigators try to explore other intense scenarios in relationship or social interactions, the experiment design may include conditions that can sabotage their relationship or lead to some serious negative impact on their emotions. Many of the longitudinal studies in relationships such as attachment style in developmental psychology very often use a field research method. Nevertheless, field research limits its information acquisitions to traditional data gathering methods. As we discussed in the beginning of this chapter section 2.2 the traditional methods rely on subjective judgement and require subjects to have high levels of self-awareness that reflect the reality.

When it comes research on social interactions and social dynamics studies, it is worthwhile to introduce a psychophysiological sensing method into field research. This project proposes and tests such methodology in real-life social scenarios. In order to achieve the goal, we also developed a platform including a smart wristband that can record physiological data (EDA and HRV) and a platform that allows a large number of subjects to be recorded simultaneously.

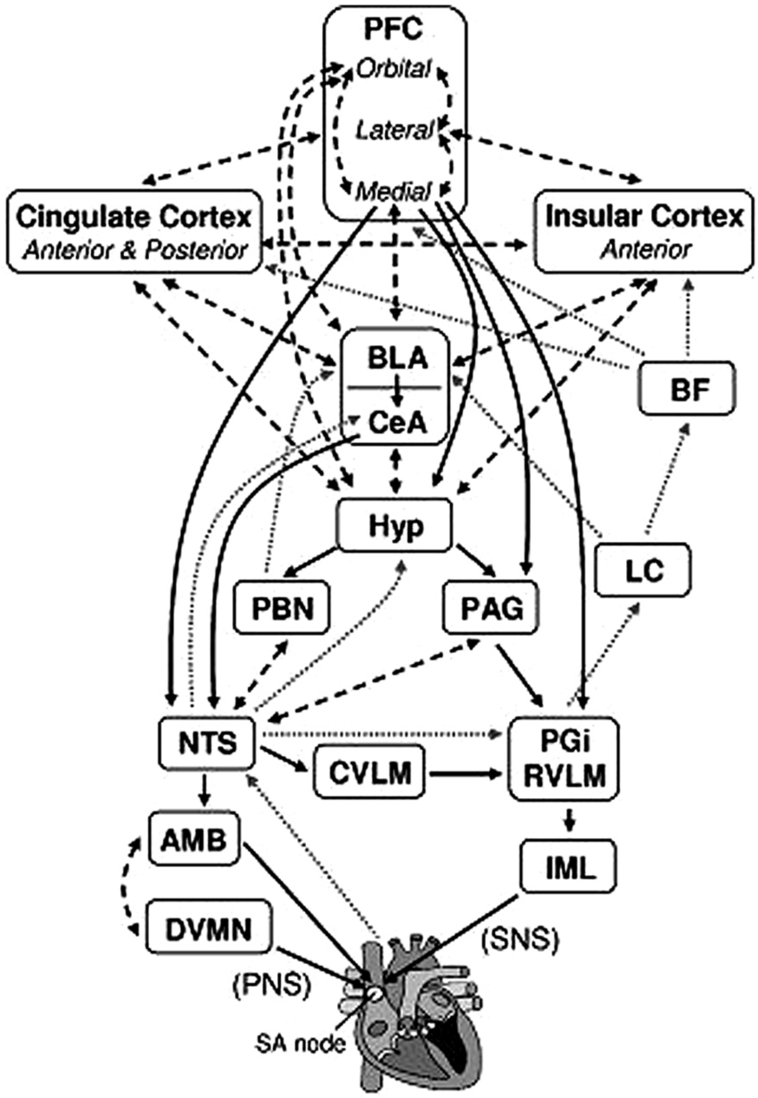

Figure 2.4 Brain structures associated with the control of heart rate [4]

Chapter 3
Methodology

In this chapter we present a possible solution to the shortcomings of social interaction studies executed in laboratory and field studies.

3.1. Problems in Social Interaction Studies

There are two main directions with which we can investigate social interactions of human beings in psychological studies. One is to observe and test in a laboratory environment where the investigators have more control over the experiment setups and can limit the variables that are unrelated to the research questions, especially for causation research. Another way is to record and observe people's behavior "in the wild". Each has its advantages and disadvantages, especially regarding the validity of the assessment. Different from scientific disciplines such as physics and chemistry, when taking human beings as the subject of the research, what we can find in a laboratory and in the real world may differ quite a bit. Regarding studies of social interactions, behaviors guided by the investigators in the laboratory are in very different social context compared to the real world. Thus, the behaviors assessed in the lab environment may be very different from how the tested participants interact in their everyday life.

3.1.1 Traditional Measuring Vs. Psychophysiology

As we discussed in chapter 2, the measuring methods are limited to the traditional methods such as interviews, questionnaires and rating from observations. These methods have the weakness of relying on subjective feedback. In order to cover that blind spot, psychologist introduced the measuring method of recording the physical feedback that the participants can hardly control as an indicator of their psychological process [1] (e.g. cognition, emotion, behaviors).

The development of psychophysiology as a branch of psychology has contributed a lot in disclosing the relationships between the mind and the body (see as Figure 2.4). Knowing what kind of physical functions indicate what kind of psychological process not only helps us understand how our body and our mind work, but also offers a new way of gathering feedback besides subjective reports from participants.

Figure 3.1 MEG Recording with a helmet-shaped neuromagnetometer and a 3-tesla Magnetic Resonance Imaging (MRI) device [5]

3.1.2 Laboratory vs. "Real-world" Environment

As we discussed in chapter 2, many relationship related studies are applying the method of psychophysiological sensing in lab environments. In the experiment design of attachment studies, usually there are few interactions between the participants. The main activities of the participants are usually following a task set up by the researchers while also passively knowing the presentation of strangers, friends, and/or romantic partners. There are also studies trying to record interactions between people [36, 91]. Taking couples studies as an example, some of

1 Definition of psychophysiology from American psychological Association (APA) dictionary `https://dictionary.apa.org/psychophysiology`

the relationship prediction studies record the physiological data while the couples were discussing challenging topics in their relationship as they were guided to [91].

The potential problems with these mentioned setups are the researchers have no control of whether the couples are actually discussing what they were told to, and even if they are, it is doubtful if they interact and respond the same way in real life scenarios. On the contrary, actually it's safer to assume the laboratory environment had impact on the participants' behaviors and interactions furthermore impact the validity of the data.

This issue raised the option to move the experimental setting into the natural everyday life environment. Running experiments in the field has been a very established method since the beginning of the social science disciplines .

Nevertheless, the procedure will not work if we simply move the laboratory sensing equipment into daily life scenarios for data recording. For one thing, people will still appear unnaturally with the wires connected and big devices covering their bodies. They will be treated differently by people around and they respond differently from how they usually behave.

3.1.3 Gaps among Disciplines

With the problems mentioned above, it is natural to come to the thought– what if we can find a way to introduce a good device that can offer accurate data comparable to laboratory equipment into field research of social interaction? With experts being very good in their own fields, there are still barriers across different fields of social science and technology. Experts from different fields are often unaware of the advancements in other fields that could be applied in their own domain of expertise. This work attempts to bridge those gaps and facilitate more interdisciplinary research by employing the latest advancements in wearable technology in the fields of social science.

With the rapid development of technology, wearable computing has been a trend and focus in the Human Computer Interaction (HCI) community. Engineers and computer scientists attempted to design wearable smart devices that can detect and understand the cognitive process of humans. Most of them faced failure due to lacking solid grand truth applications from psychological perspectives. In other words, the wearable smart devices native to the technological fields usually lack psychological methodology in their design, rendering them useless for psychological studies. The data recorded by these devices fails to represent the psychological reality and processes they aim to track. On the other hand, the social sciences, such as psychology, are often not technologically advanced and don't keep track of the new technologies and trends happening in the HCI field.

We will discusses more in detail about the weaknesses of the wearable smart sensing devices available in the market in them next chapter in section 4.2 and section 4.2.1

3.2. Possible solution: Computational Psychometrics

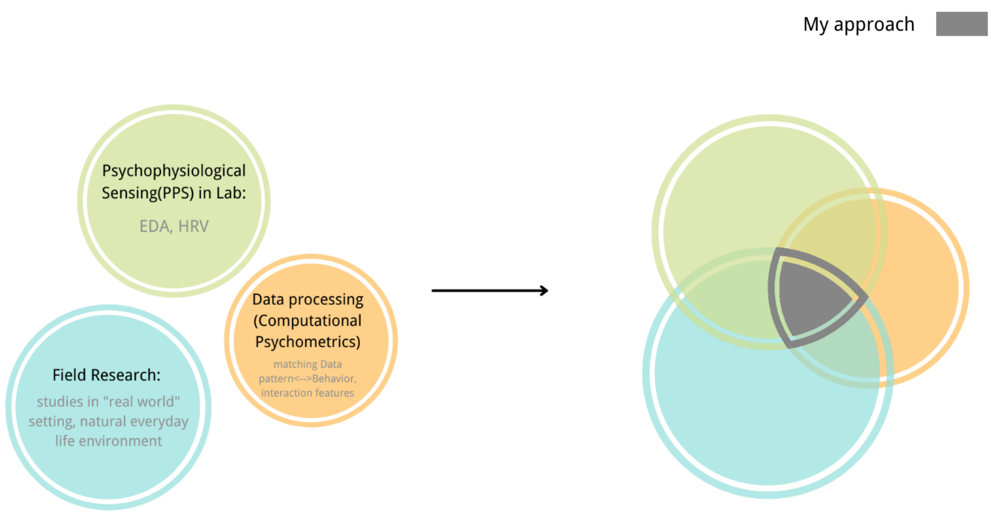

Figure 3.2 Concept Graph of the Methodology. Focus of the work is to combine the advances and methods from the three respective fields in the left side to attempt finding a new area of research where we can use wearable sensing an computational psychometrics in field research.

This work proposes a new approach of introducing a psychophysiological sensing method into "real-world" scenarios for data gathering by using custom small wearable devices that can offer comparable useful raw data (see as figure 3.2).

There is a new community established for the assessment of multi-modal analytics of performance assessment such as behaviors [92]. Mathematicians and data scientists from this community proposed the concept of computational psychometrics as a mix of "stochastic processes theory, computer science–based methods, and theory-based psychometric approaches that may aid the analyses of complex data from performance assessments, including collaborative assessments " [93,94]. This method uses computer science-based means (e.g. data mining, machine learning) to process multi-modal data for interpretation and analysis, especially for human behaviors and well-beings.

The solution we propose for social interaction studies aligns with this new field for gathering and processing psychological data.

3.3. Concept Testing

In order to test our concept and improve our experiment procedures, we have arranged a series of pilot field studies: a. to test wearable sensing in the field scenarios; b. attempt several social interaction studies; and c. to research emotion detection and computational psychometrics. Although they do not make the main body of this work, they led this project to it's current state so they are worth being brought into attention mostly as the historical context of this work. Some of these works were presented and published at multiple venues mostly related to the ACM community.

3.3.1 Field Tracking

Multi-sessions Field Recording at UbiComp 2019

2019 September, we conducted a Multi-sessions Field Recording at UbiComp 2019. The goal of this work was to attempt a large scale recording of head and eye movements of the conference attendees. As a part of a tutorial and a workshop on smart eyewear we recorded EOG and head movement data of over 80 subjects attending the conference. From the recorded data we hope to recover and categorize the activities the subjects were performing during the conference, such as talking to each other, attending talks and workshops, etc. In addition, we have gathered data on whether the participants liked or disliked the activity they were performing. By correlating these recordings with each other and the conference schedule we hope to find and quantify the data features useful for social dynamics description. Each participant was handed out a pair of Jins MEME glasses and an Android smartphone running the recording software. The smartphone application allowed users to label their current experience as "like" or "dislike". The label was stored together with the EOG and head movement data from the Jins MEME. Since the recorded data timestamps were given by the device's local time, for time synchronization purposes the application was recording the current GPS time alongside it's local time.

Originally, in addition to recording the MEME data, each smartphone was broadcasting BLE beacon packages and simultaneously listening for such packages from other devices and hardware beacons. Thus the devices could "hear" each other and using the RSSI (signal strength) could calculate approximate distance to each other and the hardware beacons placed around the venue. However, during the study it was found that the Jins MEME may disconnect from the phone and require rebooting if the host smartphone is performing too many Bluetooth operations. Thus the relative positioning recording was abandoned in order to improve the Jins MEME connection stability. The reason for this issue remains unknown, but is possibly related to the Android version and some particularities of the Bluetooth implementation on the Jins MEME side. Unfortunately, even disabling the relative positioning did not solve the issue completely. So data for many participants was not recorded or recorded inconsistently.

Although due to the technical issues the recording did not go as smoothly as planned, in total 7.75 GB of data for 80 participants was gathered. The recordings are abrupt and present a challenge for a proper analysis. Also due to the decision to stop the relative position measurements, the attempts to use this dataset for social dynamics tracking, as was intended initially, are rather futile. Several papers have demonstrated the possibility of extracting meaningful information from EOG and head movement data alone [95], so this dataset is still useful for further data-mining, but the analysis will not be presented in this work.

3.3.2 Social Interactions

The following work aims to explore the potential utilization of personal space surrounding the individual to improve the quality of social interactions. In this project we attempt to project information in an artistic form around the user in order to give the surrounding people more information.

Visualization of Mental Status

The project of Wearable Aura displays interactive projections that reflect mental state of the user on personal space to enhance communication. The design concept of it concentrates on the representation of the mental states and aims to improve inter-human communication.

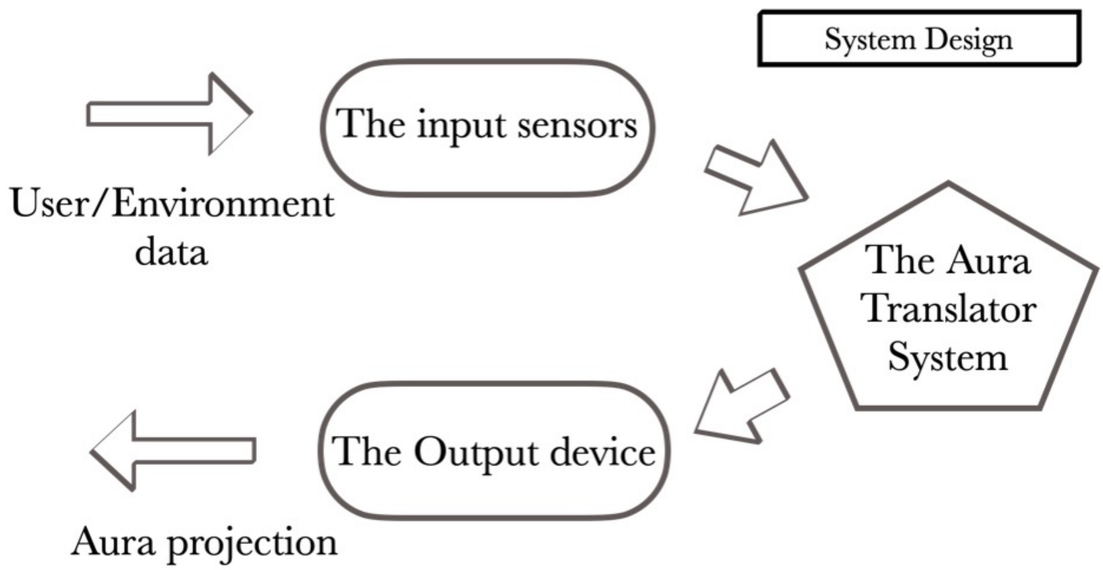

Figure 3.3 The mechanism of the system design

The Wearable Aura system requires two input groups of sensors: motion and space awareness sensors and cognitive state sensors. The motion and space awareness sensors are basically a group of distance sensors and an IMU. This group will be used to map the surroundings of the Aura in order to adjust the projection. IMU will capture the movements of the user for advanced visual effects. The information will be delivered to the Aura Translator System. The cognition input sensors will offer the information of the user's cognitive status.

The Aura Translator System processes the psychophysiological data such as EEG, user's motion and surroundings map and renders graphics reproduced by the the Aura Projector. Since the system is space-aware, it can adjust the animation movement to avoid overlapping with the surrounding objects and account for user's movements. The Aura Translator System will render the animations according to the input data. The psychophysiological data will determine the status of the aura animation; the motion data will determine the behavior of the animation; while the data of the environment will determine the position of it. The output is an animation rendered by the Aura Translator System and projected around the user.

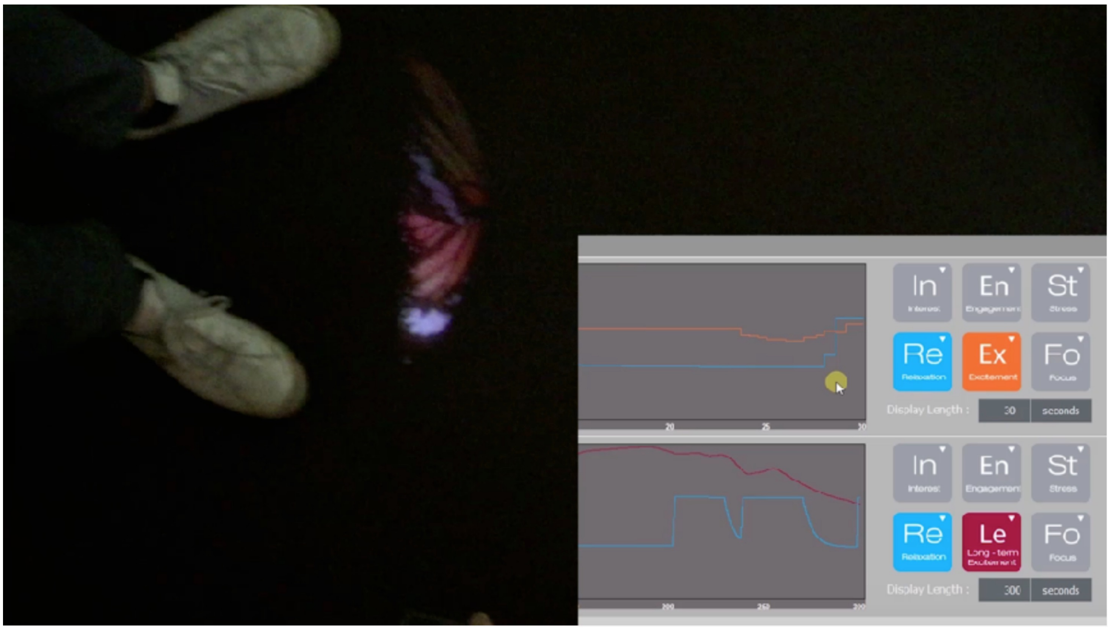

Figure 3.4 User test with Emotiv headset. The Official Software claim the Orange Line Represent Excitement Level; The Blue Line Represent the Relaxation Level.

User tests in Lab

In the lab environment user test, we used Emotiv Insight, a wearable headset, as the cognitive input sensor. Emotiv Insight has been available on the market for general release since 2015. It is a mobile EEG system. The release party claims Emotive Insight is able to detect emotions such as excitement, relaxation, boredom and so on. All this data can be displayed through an EEG software. We asked a male participant to wear the Emotive Insight, and used his excitement and relaxation data group (see in Figure 3.4) as the input data. We utilized an animation of a flying butterfly to present the aura. The flying speed indicated the excitement levels. The speed of the animation controlled by the keyboard of a PC (S:slow down 5 frames per second; Q: speed up 5 frames per second) with an Aura translator system coded through UQ. The laser projector rendered the butterfly animation through wifi connection with the PC. During the test, the participant perceived the butterfly flied faster and more aggressive when the excitement level went up and slower and calmer when excitement levels went down.

User tests on Street

We also did user tests outside of lab and walked in the street with the prototype. We tested qualitative feedback on people's reaction to Wearable Aura , with the of Wizard of Oz approach. When we were walking on street, the projection was manually controlled, and the projector was hidden from the subjects. The Aura projection was following the user and interacting with people around the user's personal distance. As of now, mid-air displays are not suitable for our use case, thus we decided to use projection for the output. For the current study, we used mini laser projectors. They are lightweight and can be worn on the body.

We ran three tests. In the first two tests, the participants were asked to wear the Aura and walk for about 10 minutes around the Hiyoshi train station. In both tests, the participants got more attention than their usual walks. In the third test, we asked the participant to walk on busy streets in the Shibuya area. The attention we got was comparable; however, the Aura wearer received a lesser degree of effective interaction. This is possibly because of the heavy traffic. Nevertheless, the participant received a long conversation with three people for more than 10 minutes including an exchange of personal information and so on. In all three tests, pedestrians who passed by the aura wearers tried different ways to interact with both aura wearers and the aura projections.

As we discussed, conceptually, depending on the user's cognitive or emotional state the pattern or image projected around the user will change to represent the current situation and behave according to user's intentions. E.g. depending on the user's desire and readiness to interact with the people around, the projection would change to a more or less inviting and welcoming one, thus aiding people in keeping a comfortable distance and reducing the stress related to social interactions. This project was published in UbiComp 2017 [96] and won the CHI 2018 Design Competition [97].

3.3.3 Physiological Sensing and Psychological Process

Below are several works focusing on cognition and physiological sensing. In this section we discuss some preliminary findings and our notes from these projects.

Cognition

In a series of Facial Thermography for Attention Tracking studies we have shown and begun to quantify a previously known phenomenon related to the fluctuation of the temperatures of the nose and the forehead related to attention and cognition related processes. [98] This phenomenon seems to be one of the most interesting and unexplored related to the subject of cognitive activity. As it is related to attention and cognitive load, it can be used to quantify the daily attention span and workload and thus the distribution of stressors throughout the day. As it is demonstrated in the referenced papers we were able to register the cognitive load related changes in the temperature of facial tissues using contactless infrared temperature sensors that are small enough to be built into the frame of daily eyewear and do not require any laboratory equipment.

Feedback Loop

The project of Physiological Signal Based Feedback Loops for Entertainment and VR is related to entertainment and aims to implement a physiological feedback loop for virtual reality gaming experiences. We are working on a VR horror game where one of the features would be the connection between player's physiological readings (such as heart rate or breathing) and game environment. For example, if we play back to the player the sound of a beating heart at a rate slightly higher than the actual player's heart rate, we hypothetically may lead player to believe that he or she is more scared that they actually are.

Among others, player's physiological signals and thermal feedback are rarely used in modern games. For this project we have developed a wearable thermal feedback system covering the player's forearms, palms and the neck. To explore the possibilities of thermal feedback we have designed the game in a way which will require the players at some point in the game to close their eyes and rely only on the thermal sensations to perform certain interactions. We also create a more immersive experience by connecting user's various physiological signals to the game environment. The game experience affects the player's emotional state and physiological readings (e.g. heart or respiration rate), which in turn changes the experience itself, forming an implicit feedback loop between the player and the game.

Many studies show the interesting and still under-explored effects of cross-modal stimulation [99]. Our senses are tightly intertwined and are constantly affecting each other, vision affects interpretation of audio, audio affects the interpretation of vibration, smell affects taste and vice versa. The whole ensemble of senses defines what we interpret as an experience. Quantification of such cross-modal effects is very difficult due to their liminal or subliminal nature, which makes projects utilizing such approaches even more interesting from both the research and entertainment perspectives. Experimental systems and experiences such as the presented one raise more scientific questions than they give answers, but a good research question is necessary.

The presented work is a VR horror game with the main focus on the thermal feedback. Cold shivers are often associated with paranormal and supernatural events. To simulate this feeling, we use thermal feedback on both hands and the neck [100]. Since the aim of this work is to demonstrate the capabilities of thermal feedback for game interfaces, the game mechanics will require the player to perform certain actions with their eyes closed, relying only on the thermal sensations on the hands.

Since virtual reality does not have to be strictly limited to the visual image, in this game the player will have to interact with invisible game entities that can be sensed as cold when the player's hand is pointed towards them. To make a more convincing demonstration that navigation in VR is possible only relying on the thermal perception without visuals, we require the player to close his or her eyes when performing the interactions (e.g. finding a spirit in the room, fighting with a spirit, etc) with the in-game supernatural entities.

In order to provide a more immersive experience we are connecting certain in-game parameters to player's physiological signals. For example, the lighting and pacing of the game could be connected to the heart or breathing rate of the player. In addition to the visuals, during the moments of tension we are replaying the sounds of a beating heart and breathing at a rate slightly higher than the players' to lead them to believe that they are more scared than they actually are. This would in turn increase the breathing and heart rate, resulting in a feedback loop. We have no scientific confirmation that such a mechanism in a VR horror environment would actually work, but hopefully it will add to the immersiveness of the experience and we could collect enough data to support or dismiss this idea.

User tests

The simplified version of the system was presented to the public during an open lab event. According to the gathered feedback the users found the heating and cooling time of the system to be surprisingly fast, and having a cold sensation for long periods of time to be very unusual and novel. Reported positive user experience with the cold sensations is one of the main reasons we make a stress on using cold sensations in this game.

We believe that connecting the physiological signals not only to the audio but to the visual game environment, such as lightning conditions may only increase the effect. The existence of such an effect and it's size is yet to be established. The project was published in CHI 2020 Game Design Competition [101] and SIGGRAPH 2020 [102].

3.3.4 Computational Psychometric: Detection and Prediction

Here we discuss several works focusing on detection and classification of certain emotional and cognitive states based on data acquired from user's physical activity.

Emotion and Haptic vibration

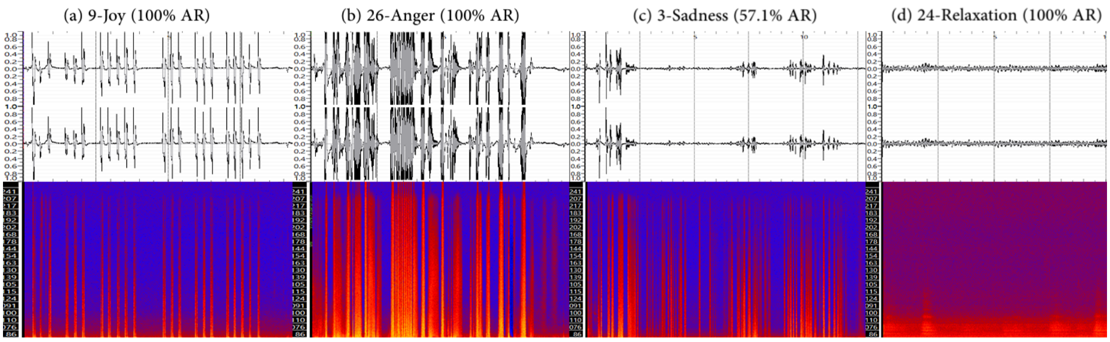

Figure 3.5 The Rhythmic Graphs of Four Emotional Vibrations from the database visualised with SonicVisualiser [6]. The top images are the waveform and the bottom images are the spectrogram. (a) 9-Joy (AR=100%), (b) 26- Anger (AR=100%), (c) 3- Sadness (AR=57.1%), (d) 24- Relaxation (AR=100%). *AR= Accuracy Rate of the sample file.*

In the project of Haptic Empathy, we found people can share emotions through simple vibrotactile feedback. We acquired 7 volunteers (3 males and 4 females) aged from 24 to 30 years old to used the TECHTILE toolkit [103] to record 28 vibration sample sets for 4 different emotions (joy, anger, sadness, relaxation). We then replayed the vibrations to the participants to test how well they could be recognized. We tested 196 trails in total. Each participant was tested with all 28 sets of different vibrations. Each emotion was tested with 49 trails.

Among 49 test trials for each emotion conveyed through vibration, there were 35 matched result for joy, 32 for anger, 25 for relaxation, and 23 for sadness (See Table 3.1). Based on that, the accuracy rate of joy was the highest 71.43%,

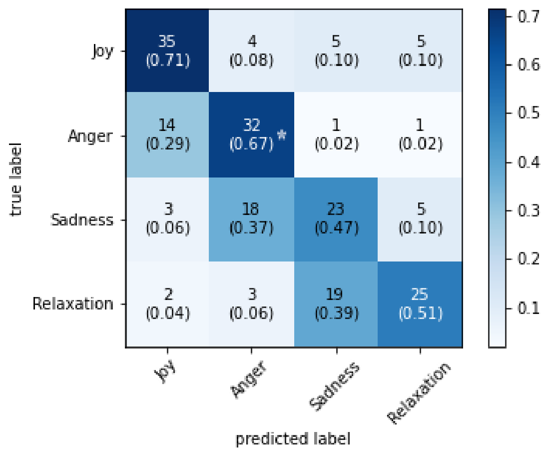

Figure 3.6 Confusion matrix for vibration patterns recognition.

followed by anger 65.30%, relaxation 51.02%, and sadness 46.94%. It is interesting to note several anomalies in the data. E.g. anger was sometimes confused with joy, however, joy was almost never confused with anger (see confusion matrix on Fig.3.6). The same holds true for anger-sadness and sadness-relaxation pairs. There was one participant who failed to choose any emotion for one of the anger vibration trials. That result was removed from the confusion matrix since it was not mistaken for any other emotions, yet remained as an incorrect result for the calculation. We can only speculate about the underlying reasons at this point. Further investigation of this phenomenon is required. For 28 tested trials, the accuracy rate for each participant ranged from 46.43% to 71.43% (Mean= 0.579; SD=0.095).

As showed in Figure 3.6, we found a very high accurate rate for high arousal emotions Joy (71%) and Anger (67%), followed with relatively high accuracy for low arousal emotions Relaxation (51%) and Sadness (47%) These results support the hypothesis that people can use vibration feedback as a medium for expressing specific subjective feelings.

We exported the recordings of the emotional vibrations to spectrograms and found that certain rhythmic characteristics were associated with specific emotions. Due to limited space, we only picked the highest recognition rate file for each emotion in Figure 3.5 as an example, yet the similar patterns are shared across the highly recognized vibration files:

Table 3.1 Mean Accuracy and Difficulty levels of the Recognition of 4-Types-Emotion Vibrotactile Emotion Samples

Emotion		**All Vibration (n= 49)**			**Others' Vibration (n= 42)**		
	Difficulty	**Correct**	**AR of All (M)**	**p-value**	**Correct**	**AR of Reg. Oths (M)**	**p-value**
Joy	2.31	35	71.4	$< .00001$ ***	31	73.8	$< .00001$***
Anger	2.67	32	65.3	$< .00001$ ***	31	73.8	$< .00001$***
Sad	2.93	23	46.9	0.0002***	17	40.5	0.0102*
Relax	2.91	25	51.0	$< .00001$ ***	22	52.4	$< .00001$***
Total	2.68	28.75	58.67	$< .00001$***	24.25	57.74	$< .00001$ ***

*$p < .05$, ***$p < .001$, *Population SD* =0.062
AR= Accuracy Rate, *M*= Mean, recognition accuracy values are in percent. *All Vibration*: all 49 tested trials. *Others' Vibration*: 42 tested trials that excluded the self input ones.

1. Joy - rhythmic patterns are mostly consistent and prominent. The rhythms tend to be more quantized than others.

2. Anger - rhythmic patterns are very prominent and constant. These patterns are more driving and abstract than joy.

3. Sadness - rhythmic content is sparse and tends to come in short patterns/gestures, with large sections of inactivity.

4. Relaxation - rhythmic content/density is quite sparse. The amount of rhythmic content is nearly non-existent.

Analyzing waveforms and spectrograms, we deduced that joy and anger share similarities in terms of rhythmic structure. These patterns tended to consist of dense clusters of rhythms interspersed with moments of inactivity. In joy, the rhythms were typically more consistent, whereas in anger the rhythms were noticeably more abstract and disjointed. Much like the relation of joy and anger, sadness and relaxation share similarities in rhythmic content. With these two emotions, the rhythmic content was quite sparse, with much of the waveforms containing large amounts of inactivity. In sadness, there were small clusters of rhythmic content. In relaxation, there were nearly no discernible rhythmic peaks. Apart from these two similar groups, we also found a connection between joy and relaxation in terms of rhythmic consistency. For example, the waveforms shown for joy and relaxation show that the activity is occurring in a mostly repetitive or consistent manner. On the other hand, anger and sadness were quite the opposite. These waveforms show rather inconsistent and varied rhythmic content.

As showed in figure 3.5, we found there seems to be a correlation between frequency signal patterns and emotions (psychological process). However, in this study we did not try to train an algorithm to detect psychological process from the features of signal data. Later on, as we will discuss soon, in both a handwriting study and frisson study, we trained a machine learning model for such kind of prediction. This project is published in CHI2021 [104].

Emotion and handwriting Features

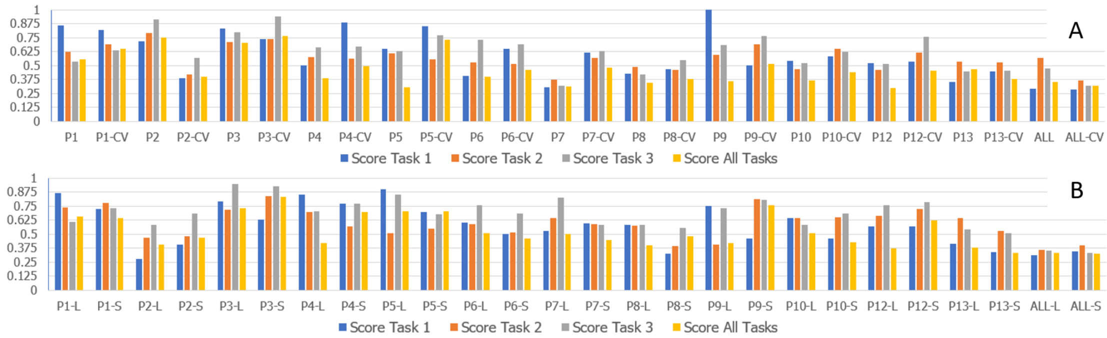

Figure 3.7 Classification scores for user-dependent model. A. Test scores sing all strokes, CV - cross validation scores for each participant's model. B. Models using only short (S) and long (L) strokes. "Score All tasks" is the task-independent classifier, "ALL" represents data for user-independent classifier.

The project of Recognizing Emotional Context Based on Handwriting Features took a rather different approach to the mental state assessment. Instead of registering and quantifying certain physiological signals, it relates upon the manifestations of the Autonomous Nervous System (ANS) activity in fine motor tasks, namely handwriting. Later using machine learning approach we managed to achieve the classification precision (1 out of 4 emotional conditions) of 66% for each individual stroke.

In order to build the classifier we used the valence and arousal scores of the SAM test as labels. The scores were grouped into "high" (6-9 points) and "low" (1-4 points) categories. The data scored at 5 points of valence or arousal was

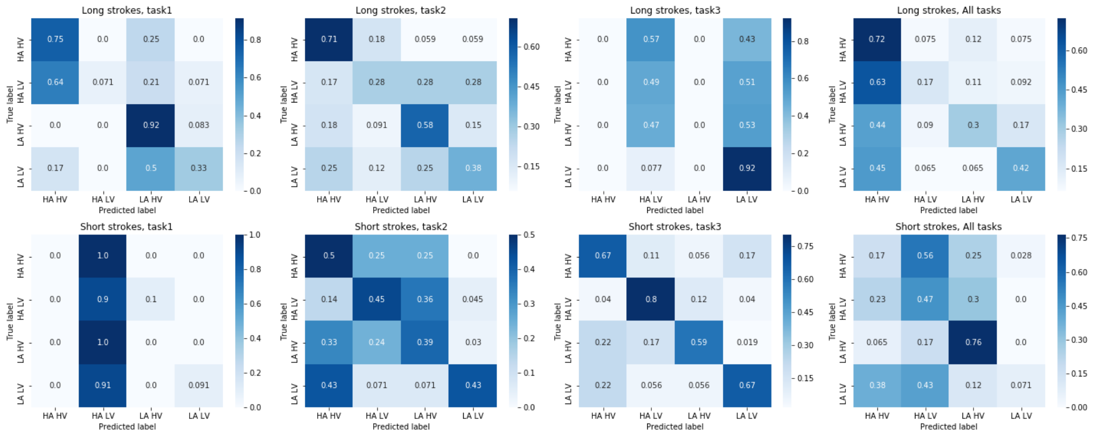

Figure 3.8 Confusion matrices for user-independent model. The stroke data was split into short (bottom) and long (top) strokes before classification.

omitted. This assures that the actual arousal and valence levels of the participant correspond to the handwriting, since not all the participants had intended emotional response to the video stimuli.

The confusion matrices of each condition are presented on Fig.3.8, and the score results of this classification are presented on Fig.3.7. We ran classification on the data sets from each task and on all three tasks combined. It was found that the classification precision changes greatly if we use only short or only long strokes. The stroke data was split into short and long strokes in relation to the median for each data set used. Short strokes gave particularly good results for task 3, as it required participants to draw mostly short lines. Accuracy for this test is reaching 66% for 1 in 4 groups classification. Using long strokes showed better results for tasks 1 and 2 with accuracy of 50 and 51% respectively. Surprisingly, the accuracy of classification of the short stroke data set for all 3 tasks was higher than for long stroke data set (47% for short and 35% for long).

On the downside, although this approach provides a reliable labeling, we have to exclude all the data that did not fit into one of the four labeled groups: 1. High Arousal and Valence. 2. High Arousal low Valence. 3. Low Arousal and high Valence. 4. Low Arousal and Valence. After applying this labeling to the stroke data, the data for some of the participants was not sufficient to build a reliable classifier. Thus this approach was used only for the user-independent model that was trained on the data from all the users.

Since fast and precise motions characteristic of handwriting are not under our conscious control, there are several mechanisms that could affect them. This work was published at AHs2019 [105].

Frisson and EDA, HRV Features

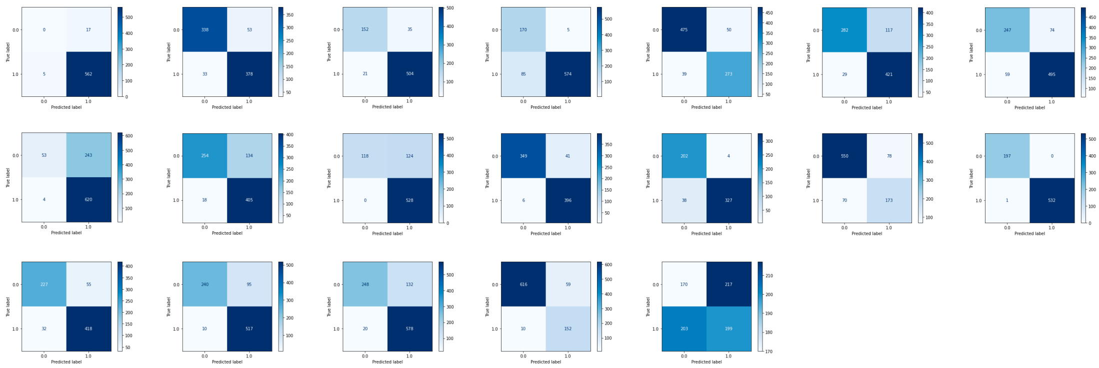

Figure 3.9 Frisson: Confusion matrix for the prediction model with LOPO-CV. A row represents an instance of the actual classes whereas a column represents an instance of the predicted classes.

In the project of Frisson Waves, we trained our algorithm model to detect frisson/shivers/goosebumps in real-time.

Participants

We recruited 33 volunteers (female= 17; male = 16) aged 22-37 years old (average 26.09) for the 30 minutes study. Once the study began, participants filled out the demographic questionnaire followed by the explanation of frisson while the investigators helped them put on the wristband with a frisson-report button and the neckband with two thermal modules placed on the back of their neck. The investigators then explained the definition of frisson and when to press the button.

Procedures

The same Thermo-Haptic Neckband and Wristband were used with a counterbalanced order of 3 parts, including part A, part B, and part C. Part A is a five minute excerpt from Gustav Holst's "The Planets: Jupiter, the Bringer of Jollity", approximately 4:00-9:00. According to previous studies, this particular 5 minutes is the part most successful in giving the chills [106]. Part B is a three minute cold thermal feedback stimulus session through the neckband with no music stimulation. The cold feedback onset period is 8 seconds: 3 seconds cold feedback "on" and 5 seconds cold feedback "off". Part C is a 5 minute piano recording from Frédéric Chopin's "Prelude, Op. 28, No. 15". It was recorded from the first live concert we held and rated most likely to have frisson from the audience. In this study, we only turn on the Thermo-Haptic Neckband in one music part throughout three sessions: A and B or B and C.

We use a support vector machine classification (SVM) algorithm in our frisson detection referring to some previous works using physiological data to detect physical or mental phenomenon [107, 108]. Our model was trained using the features extracted from a sliding window of one minute. The window was labeled as a frisson event if the button was pressed within the window. We extracted four EDA features and three HRV features using Neurokit2 [109] from raw signal.

To achieve a better generalization of the model, we normalized the features and applied leave one participant out cross validation (LOPO-CV) to divide data into training and testing sets. The classifier with the best performance presented an average accuracy score of 85.78% (sd = 11.23%) with an average precision score of 81.75% (sd = 12.48%). Figure. 3.9 illustrates the overall performance of the model.

3.3.5 Lessons Learned

These previous works and pilot tests offered valuable insights into how we can improve the details of the setup and contributed to the presented vision. In those studies we could get the experience necessary to conceptualize the requirements for the experimental setups, develop the details of the approach and the tool-set presented in the following chapters.

Although none of these projects cover all of the aspects of the developed framework, when combined, they address every nuance we need to pay close attention to: difficulties associated with the large scale field recording at Ubicomp 2019; social interactions in Aura project, emotional context in the handwriting studies.

Field Setting

From the Ubicomp 2019 recording, we found that for such a field study, people were very often in a rush when we tried to recruit them to participate in the experiment. We were waiting for the potential participants in front of the door of the lecture room before the speech started. Most of the time we only had one or two minutes to explain our wish to have them join this study. It was critical to contact the conference organizers and get approval for conducting such studies with the attendees of this conference, and it helped us to shorten the process of explanations. Most of the people who joined as subjects had at least basic understanding of experiment procedure. Nevertheless, it is still very important to make every setting as simple as possible in order to encourage the participants to volunteer for the study as well as to understand what are they expected to do from their side during their participation. Especially in our case of the UbiComp recording we had no access of monitoring the devices once the set left us to the participants. This also turned out to not be a good idea. We will discuss the reasons a bit more later in 3.3.5.

Data Processing

While the recording itself in the UbiComp 2019 study was a huge success, we faced quite some obstacles for data analysis. First of all such kind of data set is very new and there is not much established ground truth for us to label and to classify the data patterns with behavior features. We had very limited information of the field reality when trying to compare the data. For one thing, we did not have the access of the video recording yet to compare what happened in the field while we see those data patterns from the recording of the participants. Another issue was the lack of a reliable clock to synchronize all the recordings.

Nevertheless, this was a valuable experience that allowed us to deduce the requirements for the recorded data and a good attempt to explore various data analysis methods related to feature extraction, classification, machine learning and signal processing.

Devices and platform

We tried many devices for field recording. One of the big issues we faced was the connection between the wearable devices and the software platforms. For instance, we had more than 50 smart phones paired up with the same number of smart eyewears. The synchronization time with each set of devices are slightly different depending on the phone's time. As it was mentioned, in a conference setting the Bluetooth connection did not prove itself as a reliable option. In addition to this we had no monitoring system to know in real time whether the data was being recorded and whether there were any connection problems.

In the Aura project we tried to use Emotiv as the input device to decode the cognitive process. Yet, the headset was not a device suitable for daily life use. We could not wear it on the street without changing people's attention focus from the Aura projection to the cyber-headset. More importantly the EEG data plot is seemingly rather noisy and the calculated scores seem to be rather random and we have no access to the raw data. More details regarding the weaknesses of the available devices such as Emotiv in the market will be discussed in chapter 4 section 4.2.1

3.4. Summary of the Concept Design

In order to explore if there are any physiological patterns connected to the exhibited behaviors and interactions between people, we use wearable devices with multiple sensors to track the participants physiological responses in several different social events. My previous works and other studies show an association between psychophysiological and behavioral responses with mental states such as stress and emotion. However, to demonstrate the viability of the proposed approach, many more studies and recordings are needed.

A number of cognitive and perceptive psychological studies have proven human perceptions and judgements based on self-awareness can be far deviated from presented facts. Unlike experiments of physics and chemistry that are well respected as basic science studies, modern social and psychological research has been struggling for more than a century to push the experiment designs and the evaluation methods less reliant on subjective data, which happened to be one of the main sources of their data, e.g. self-report feedback, interviews which can be subjective from both interviewee's and interviewer's sides, observers' rating of videotaped behaviors, etc. As mentioned above, physiological responses can barely be consciously controlled compared to self-reports and interviews and therefore may reveal different insights and patterns than those captured through traditional measures.

We tested our method in a series of studies from different perspectives. We adopted the successful practices and modified the aspects that need improvement for further studies.

Chapter 4
Platform

In order to test the new approach we discussed in Chapter 3, in this chapter we present the conceptual design aspect of the platform that can gather useful data which delivers meaningful results.

4.1. Requirements for the System

The proposed approach relies on our psychophysiological sensing platform that consists of a wearable smart wristband, control software and several jupyter notebooks for analysis. The wristband includes three major sensors on the main board (see as figure 4.11), EDA, BVP and accelerometer. The accuracy of this device has been acknowledged by the HCI communities as it has been used in a couple of publications for international conferences (TEI, AHs, World Haptics) [110–112]. It is also currently used in medical school research to track doctors' psychophysiological sensing during practice.

The purpose of developing a brand new device was the lack of comparable devices on the market. Since such a project requires recordings of large amounts of raw unaltered data, there is no cost efficient solution that could be used. In order to be usable for this project the tools have to match certain requirements.

- **Cost efficiency.** Since it is supposed to be worn by large number of untrained participants, the devices have to be cheap and robust.
- **Access to Raw Data.** Data produced by the platform has to be unaltered by any algorithms or filtering. If it is altered, the algorithms have to be transparent, otherwise it may add complications to the data analysis.
- **Simultaneous Recording.** The platform has to be able to function with

tens or hundreds of simultaneously active recording devices. This is a serious limitation for most of the devices available on the market, as they are designed to work with one host device (smartphone or a laptop). Using multiple host devices makes it difficult to synchronize the data and assure that the recording can be analysed as a whole and correctly depict the synchronous physiological responses to certain events and interactions.

- **Build and placement.** The sensing hardware must not be distractive or cause inconvenience to the user. E.g. a large MEG, fMRI, EEG setup, even wireless would distract if not the subject, then other study participants interacting with the subject. Another aspect is the fragility and the sense of safety. The participants should not be concerned about injuring themselves or damage the device. Which implies certain build quality and protection of potentially fragile or dangerous parts of the device, giving users a sense of robustness.

Commercially available devices are either expensive, or are not suitable for massive simultaneous recordings, or do not provide the unaltered data, in the attempt to make the recordings look better than they are (excessive filtering, no transparent data processing pipelines). With this in mind there was a need to develop our own software/hardware platform.

4.2. Sensing Devices Available on the Market

This section we will discuss some of the devices currently available on the market from the perspective of their potential usage within the proposed approach.

Consumer Grade Fitness Trackers

Consumer grade trackers such as Fitbit or iWatch do not give sufficient access to remotely usable data and are mostly limited to heart rate tracking. Any alterations to the existing capabilities are nearly impossible and syncronyous recordings are difficult to implement.

Pupil Labs Eye Trackers

Figure 4.1 Pupil Core Headset from Pupil labs website [7]

Pupil Labs provides a very reliable and affordable eye tracking platform (See Fig.4.1). It offers lightweight (23 g) with two eye facing cameras on the sides that are capable of 200 FPS video streaming, which allows to track the eye motion and relative pupil size at 200 SPS. Eye tracking cameras work in the IR spectrum and are equipped with IR LEDs for reliable eye-tracking in low-light conditions. Devices come with open-source software platform allowing flexible data management and processing and supporting plugin functionality. Although it is an outstanding device with very flexible software there are certain issues that prohibit its application for this project.

- **Cost efficiency.** Although it is very affordable for the functionality it provides, the price tag of 2000-3000 EUR is rather high for large scale recordings. In addition to the price, each device has to be connected to a host computer via a USB cable, which increases the cost of the setup and limits user's mobility.

- **Access to Raw Data.** Data processing is done in software, and access to all necessary data is provided. Processing algorithms are public, as the software is open-source.

- **Simultaneous Recording.** Although simultaneous recordings of multiple subjects on the same host machine are not supported out of the box, if the computer hardware allows it, it could be possible to run two instances of the recording software. Also it is possible to use several machines. Although synchronization of recordings is not supported, it is possible to implement this functionality using the plugin mechanism.

- **Build and placement.** Since the device is placed on the face, it may cause some unease or be distracting, changing the natural flow of social interactions. Another issue is that it requires a wired connection, limiting the subject's activities. Although the Pupil Labs engineers did a tremendous amount of work to make the device as lightweight as possible, it resulted in the device feeling somewhat flimsy. This together with the price of the device makes it hard to not pay attention to the device and how it is used.

Empatica E4 Wristband

We used the Empatica E4 wristband as an example of a wearable device available on the market, and used it to collect physiological data (Figure 4.2). It comes in a small wearable form-factor, is safe, lightweight and easy to use. The device is controlled with a phone app. The device looked very promising, so we ran some feasibility tests with it. During the E4 Wristband tests we had encountered numerous issues with noisy data and poor electrode contact with the skin, and during the study we had to ask the participants not to move their hands to avoid noise in the data, which clearly limits the potential applications of this device. Another issue was the loss of all the data recorded with the android app. Thus we had to use an iOS device.

For measurement of the heart rate it relies on a photoplethysmograph measuring the Blood Volume Pulse (BVP)4.2. It measures the change of the amount light reflected from the skin in a narrow spectrum that is caused by the increase/decrease of the amount of blood on each heartbeat. Modern BVP sensors can be small

Figure 4.2 Empatica E4 Wristband and Technical Specifications from Empatica website [8]

enough to be built in smart-watches and sensitive enough to clearly detect each heartbeat, which allows us to extract the heart rate variability from the recorded data.

This device is also equipped with a thermometer monitoring the user's wrist skin temperature, but in this work we did not use this data as there is no known evidence of correlation between wrist temperature and emotional state.

E4 is also equipped with 2 EDA electrodes placed on the inner part of the wrist. Currently E4 is almost unique in it's capability to measure EDA. Although the placement of the electrodes is suboptimal, as it is preferable to place the electrodes on the glabrous skin of the fingers. However, technically it is close to impossible to produce such a device in form of a ring, as the electrodes have to be galvanically interconnected and be based on the same reference point.

- **Cost efficiency.** Each E4 costs around 1700 EUR, which is rather expensive for large scale simultaneous recordings. The cost of parts of this device is incomparably lower to its price tag.

- **Access to Raw Data.** Data is accessible only through the Empatica cloud service, which raises concern about the privacy. Also there is no access to raw sensor readings and the data provided by the Empatica platform is obviously somehow processed.
- **Simultaneous Recording.** Since the devices are supposed to be used with a phone, it seems rather hard to manage multiple simultaneous recordings. Although this set up is sub-optimal, finding a reasonable solution presents a challenge, but there may be a way to solve it in software.
- **Build and placement.** The device is well built and does not cause any significant amount of discomfort. Since all the sensors are embedded in the wristband, it feels robust and does not require any attention.

Shimmer

iMotions Shimmer is very similar to E4 in it's capabilities and form factor. However it does provide better quality data and software compared to E4.

- **Cost Efficiency.** Cost of over 500 EUR is rather high, but much more welcoming than that of the E4.
- **Access to Raw Data.** We did not acquire a unit to run thorough tests, but at a glance it seems to be providing reasonably good output.
- **Simultaneous Recording.** Manufacturer's software poses severe limitations on the hardware that shall be used not providing enough flexibility, however the simultaneous recordings seem to be reasonably well implemented.
- **Build and Placement.** Although it is a little larger than E4, it is still small enough not to cause additional distractions. One problem may be the use of 3 fingers for sensor placement, as that may cause some degree of inconvenience.

DIY kits:

Emotibits, Mikroe click, Sparkfun Biometrics, and other DIY physiological sensing kits and boards.

Recently there are a myriad of DIY kits and hobby project boards on the market that are equipped with various biometric and physiological activity sensors. For the sake of simplicity we will group them together as most of their aspects, pros and cons, are rather similar.

The quality of the recordings directly depends on the sensing ICs and circuitry used in the module and may vary greatly from one implementation to another, from one supplier to another. In most cases the hardware is supplied with basic recording software or just libraries to be incorporated into other projects.

- **Cost Efficiency.** In the vast majority of the cases the price of such modules is very low. However, there are additional costs related to assembly, enclosures, etc.
- **Access to Raw Data.** Most of the modules provide direct access to the sensing ICs, which makes the data as raw as it could possibly be.
- **Simultaneous Recording.** Software has to be developed from scratch in most cases. On the positive side this allows to design the platform to be as flexible as necessary.
- **Build and Placement.** Such devices require a custom built case. If used as is, the exposed circuitry and wiring is bound to distract people. On the other hand, if designed well, such a device could be small and suitable for this project.

4.2.1 Limitations

With the above in mind we could not select one platform to be used as all of them had significant drawbacks, such as prohibitive costs, lack of flexibility on the software end, support for the necessary sensing modalities, etc. Thus it was decided to develop our own platform that would be cheap, easy to use and be capable of performing large scale recordings with multiple subjects simultaneously.

Table 4.1 A Brief Review of Wearable Devices Available

Device	Sensing	Cost	Raw Data	Simultaneous Recording	Position and Fitting
Fitness Trackers	No useful data	/	/	/	/
Pupil Headset	Eye tracking, pupil dilation	Expensive: 2000-3000 USD	Provided	Not supported	Face: unease or distractive
Empatica E4 Wristband	EDA, HRV, etc.	Expensive: around 1700 USD	No	Not supported	Wrist: comfortable
EMOTIV Headset	EEG	Expensive: 850 USD	No	Not supported	Head: unease or distractive
DIY kits	Various kinds	Low	Possible	Has to be developed	Depends on the self-build design

Our aim was a device akin to Shimmer, with it's optimal sensor placement, user-friendly design, and wireless autonomous functionality, but with a price tag and flexibility like the various DIY physiological sensing kits.

Before implementing such a platform we have performed a series of feasibility tests and replication experiments of the works presented in the Chapter 2. Below is a brief description of the performed tests followed by functional description of the designed software/hardware platform.

4.3. Sensing and Emotion Tests

The tests were conducted with four participants (one male and three female). The ages participants ranged from 24 to 27. Participants were KMD students and KMD alumna. Participants signed the informed consent form and were allowed to stop the study whenever they wish. They were briefed with the goal of the research after attending the tests to avoid any biased behavior. Any discomfort during the test should be reported and participants were free to stop the experiment at any time. Each participant was personally informed that he or she was going to see many pictures of spiders and snakes among others. The participants were asked to read and sign through the consent before the experiment. They gave their written informed consent and had normal or corrected to normal vision.

4.3.1 Procedure

In this series of tests participants were asked to view a series of pictures from the Geneva affective picture database (GAPED) [113]. The pictures have been rated on valence and arousal scales and grouped into positive, neutral and negative categories according to the valence rating by the database maintainers.

The Geneva affective picture database (GAPED is a 730-picture database focusing on valence and normative significance. Four specific negative contents were: spiders, snakes, and scenes that induce emotions related to the violation of moral and legal norms. Positive and neutral pictures were also included: Positive pictures represent mainly human and animal babies as well as nature sceneries, whereas neutral pictures mainly depict inanimate objects.

For the tests we picked 30 pictures with the highest and lowest arousal ratings from each out of 3 groups. In total having 6 groups of pictures:

1. High Valence - High Arousal.
2. High Valence - Low Arousal.
3. Neutral Valence - High Arousal.
4. Neutral Valence - Low Arousal.
5. Low Valence - High Arousal.
6. Low Valence - Low Arousal.

The test consisted of 6 sections. In each section 30 pictures from each set were presented to each participant in the form of a slide show. Each picture was on the screen for 6 seconds followed by 4 seconds of blank white screen in between the pictures. Participants were asked just to watch the pictures while their physiological data was recorded. Experimenters did not talk to the participants during the session to avoid biasing the data. The participants might request a break in between sessions if needed. At the end of the experiment, the participants were thanked and informed with more details about the purpose of this study. Recordings and discussion are presented below.

Following are the goals of this test:

1. Determine which of the physiological responses would be most appropriate for the actual testing,

2. See if there are any correlations between physiological signals and emotions in this small sample size,

3. Practice a means of organizing the test as well as collecting and reviewing data, and finally,

4. Explore wearable devices such as E4 consistent with other laboratory devices.

4.3.2 Electrodermal activity (EDA)

As mention before, skin conductance refers to the variation of the electrical properties of the skin in response to sweat secretion [69]. In the pre-test, we used AD Instruments PowerLab 16/35 and GSR amp set (Figure 4.3) and E4 wristband to measure skin conductance. In the comparison of weather E4 is good enough to collect skin conductance data, we used the Skin Conductance data collected from PowerLab 16/35 and GSR amp set as ground truth assessment.

Figure 4.3 DAQ system Powerlab 16/35 with GSR amp device [9]

In the pre-test, we only utilized one channel of PowerLab 16/35 connect with GSR amp kit to measure skin conductance. GSR amp come with GSR Finger Electrodes: bipolar finger electrodes supplied with the GSR Amp; finger plates are made from brightly polished stainless steel and are fitted with Velcro tape.

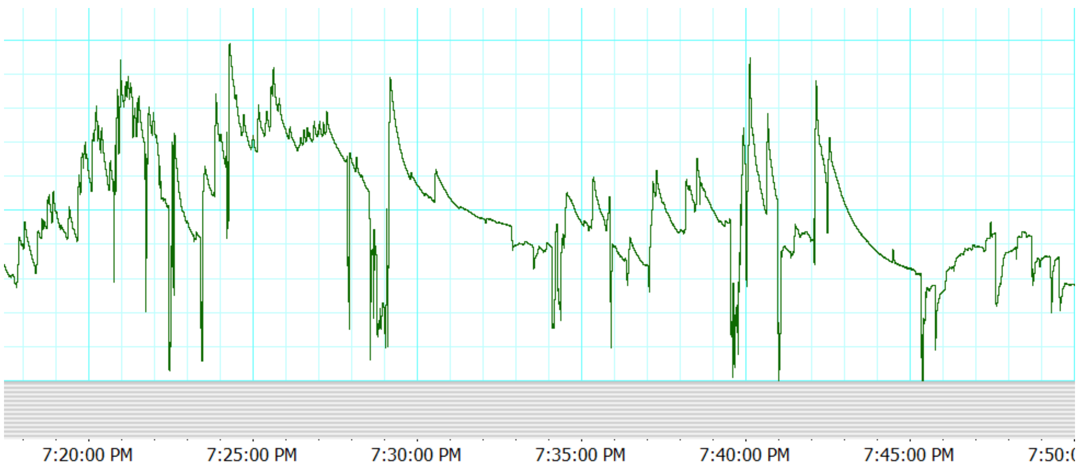

Figure 4.4 GSR graph of 6 sessions: Negative High - Negative Low; Positive High - Positive Low; Neutral High - Neutral Low

4.3.3 Heart Rate

Figure 4.5 Fingerer Pulse Oximetry

In this pre-test we used a pulse oximeter to collect heart rate and heart rate variability (HRV) data. Pulse oximeters are medical devices that indirectly monitor the oxygen saturation of a patient's blood (as opposed to measuring oxygen saturation directly through a blood sample) and changes in blood volume in the skin, producing a photoplethysmogram. However, in medicine, one's pulse represents the tactile arterial palpation of the heartbeat by trained fingertips Figure 4.5. In this pre-test, we asked all the participants to put their index finger into a pulse oximeter. We found that heart rate can raise when the participants view high arousal pictures comparing low arousal pictures (Figure 4.8 Blue graph).

4.3.4 Pupil Dilation

The use of pupillary response, or pupilometry, in psychological studies dates back to 1960 when Hess and Polt described the link between viewing emotionally toned or interesting visual stimuli and changes in pupil diameter [114, 115]. This work demonstrated that certain emotional and/or cognitive processes are reflected in the pupil diameter. Later on it was demonstrated that the pupil diameter also responds to emotional arousal, cognitive load and memory tasks [114, 116, 117].

Pupil responses were observed even in infants exposed to certain stimuli [118], which makes it a valuable tool for pre-verbal or non-verbal studies.

Recent studies show a tight connection between pupillary response and locus coeruleus, the noradrenergic hub of the brain, that is tightly connected to the prefrontal cortex among other circuits. Locus coeruleus is activated by stress and pupillary response related to its activation can serve as an indicator of cognitive and emotional processes [115, 119, 120].

Henderson, Bradley and Lang's emotional imagery and pupil diameter study suggests that the pupil's response during affective picture viewing reflects emotional arousal associated with increased sympathetic nervous system activity. However, the valence level and arousal level of the offered imagined scenes are not clear. In another study participants had to imagine scenes describing pleasant, unpleasant, or neutral events. Recordings of the pupil diameter show a significant effect of the pleasant scenes. Since the emotional imagery is widely used in clinical assessment and treatment, researchers suggest the use of pupil diameter as an index of emotional engagement in clinical usage. [121, 122]

We used an eye tracking device called Pupil Headsets from Pupil Labs to collect pupil dilation data. Pupil diameter was continuously sampled at 100Hz using Pupil Headsets. There are three cameras attached to the frame akin to a pair glasses. The front viewing camera records the field of view of the participant. Each side camera records and streams the closeup of the eyes of the participant. The side cameras are used to estimate where the participant is looking in 3d (vergence) and calculate robust binocular eye movement data. Figure 4.1. Later this data is mapped on the front-viewing camera video in Pupil Labs software. Pupil Labs offers two software kits for the pupil headsets: Pupil Capture and Pupil Player. Pupil Capture receives video streams from the tracker and performs the processing: detects user's pupil, tracks gaze, tracks markers in environment, streams data in real-time over the network, and is capable of recording and saving data in an open format. Pupil Player is a media and data visualization tool at its core. It can be used to look at Pupil Capture recordings by visualizing the data and exporting it to other formats.

Before the experiment, both of the side binocular cameras need to be adjusted so that they can have a clear view of both eyes as well as the center camera so that it can record the field of view of the participants. When adjusting cameras, the experimenter reset the 3D eye models in the Pupil Capture software to fit each participant until the confidence level of pupil tracking was close to 1.0 for each eye.

Then the experimenter helped participants to calibrate both eyes for eye tracking by following the instructions from Pupil Capture software until the red dot (Figure 4.7) on the investigator's computer screen matched to participants eye gaze.

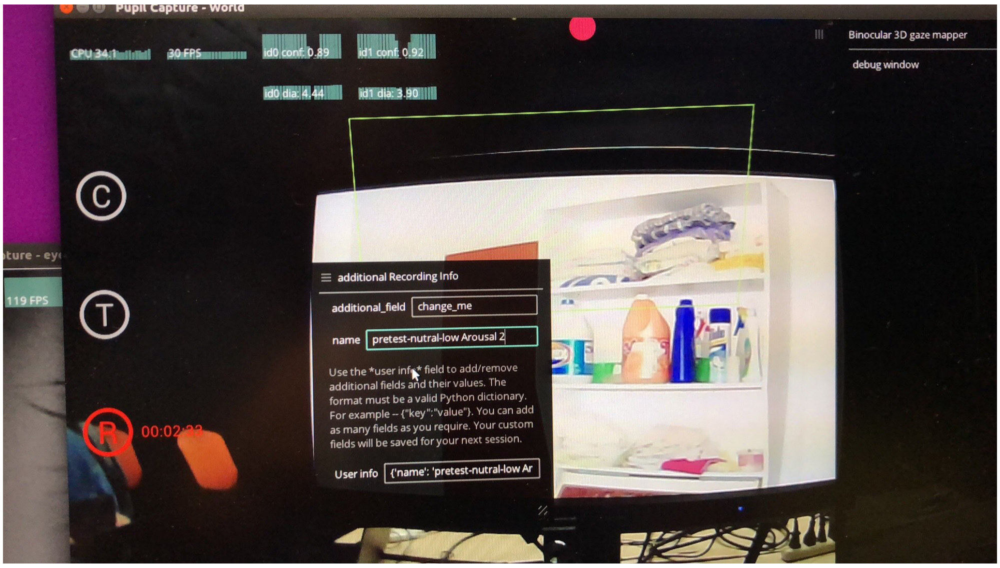

Figure 4.6 Pupil Capture Recording and Eye tracking from Pupil Lab

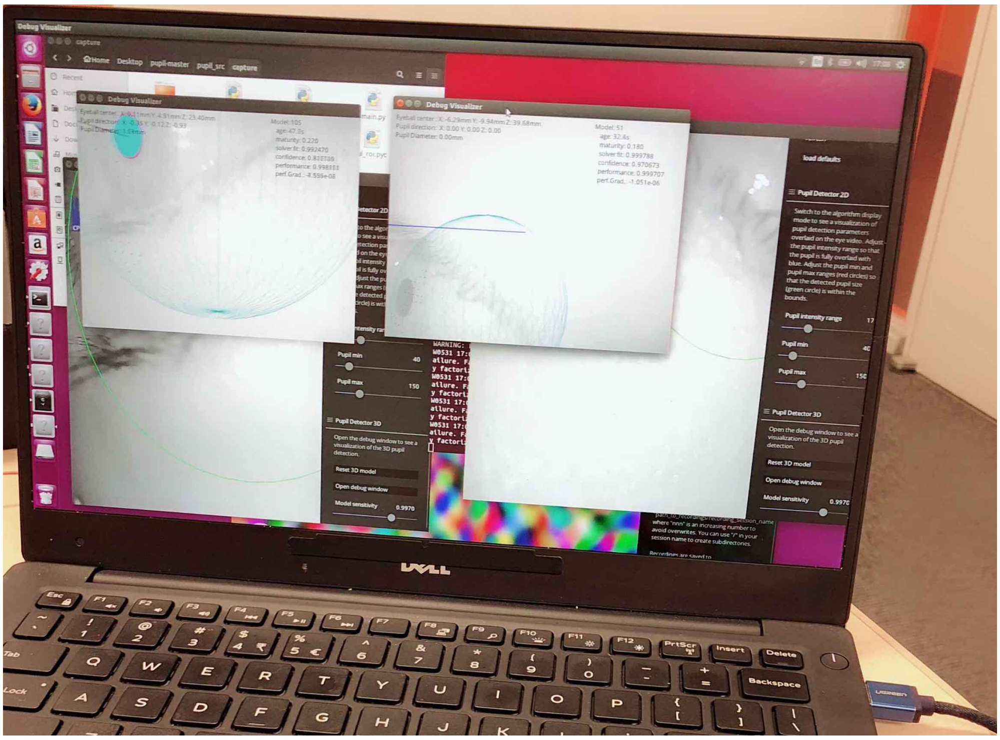

Figure 4.7 Pupil Capture Pupil detecting and eye gazes from Pupil Lab

In this test we observer the pupil dilation response to be related to the reported arousal levels and the arousal ratings of the presented pictures. In order to have more data, we kept pupil headsets for the next step of the experiments.

We use thermal camera to collect facial temperature. We try to calculate the temperature differences between nose and forehead.

In the pre-test, we did not find any consistent temperature differences between nose and forehead throughout presenting different groups of pictures. It seems that the temperature changes are mostly caused by increase in cognitive load and concentration rather than emotional stimuli.

4.3.5 E4 comparison with Lab devices

We have tested a couple of wearable devices and compared the data with the laboratory equipment. Taking E4 as an example, we will discuss how we compared the physiological data we gathered from wearable sensing device such as E4 and

Lab devices such as PowerLab. Although E4 is far from optimal for this project we have conducted a series of tests. It is rather doubtful that the physiological signals collected by E4 can be used to detect emotions.

In the study described below we use medical equipment as the baseline, but all the recordings are replicated using a wearable device. The motivation for this decision is to compare and demonstrate that the sensitivity of wearable sensors available on the market today is sufficient to replicate the lab studies discussed above, which would imply the possibility of utilization of these devices in the emotion tracking application scenario.

1. Electrodermal activity sensor for skin conductance (EDA),
2. Photoplethysmography sensor for continuous heart rate and heart rate variability measurements.

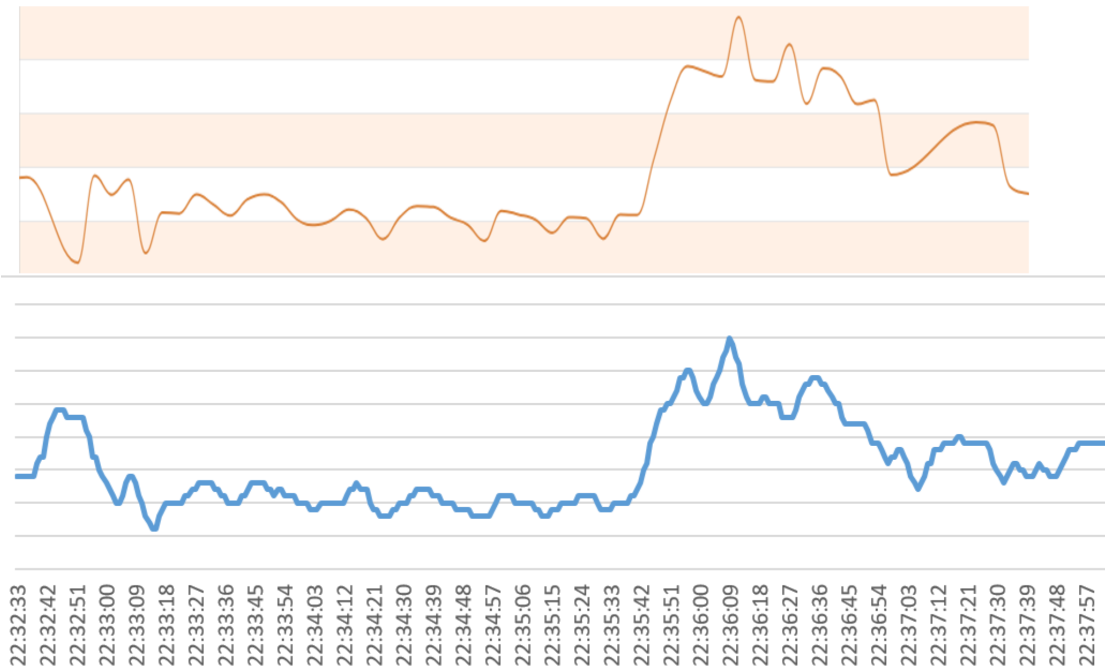

Figure 4.8 Comparison of HR E4 Wristband (orange line) and Finger Oximeter (blue line)

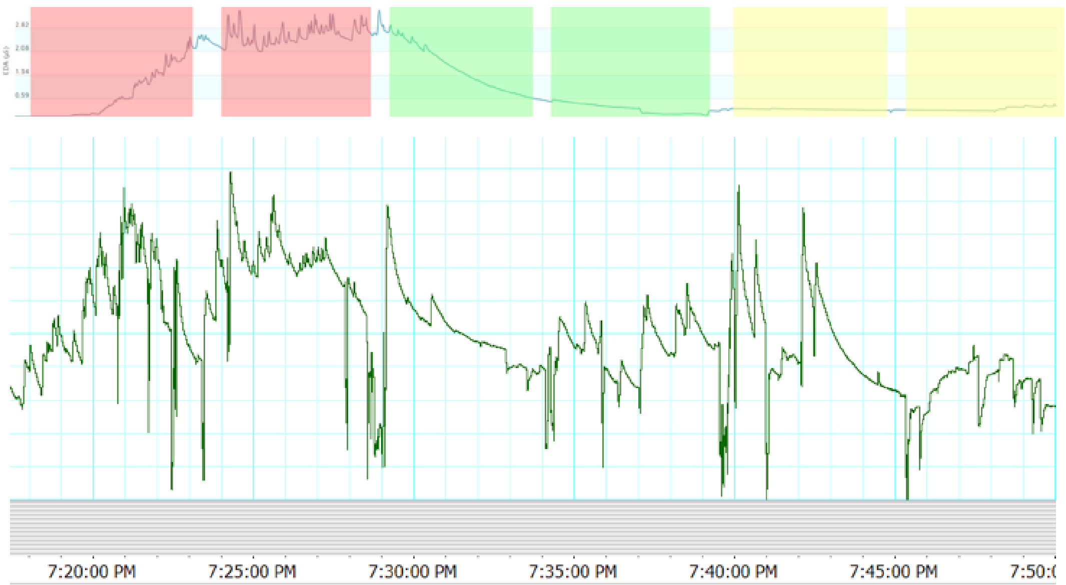

Figure 4.9 (Top) EDA signals of E4; (Bottom) from PowerLab 16/35. Six Emotional Sessions: Red –negative high arousal, Red–negative low arousal, Green–positive high arousal, Green–positive low arousal, Yellow–neutral high arousal, and Yellow– neutral low arousal

We compared skin conductance, HR, and skin temperature collected from E4 with accurate equipment usually applied in Lab environment . From Figure 4.8, we can see that the heart rate graph of E4 Wristband is not ideal, but close enough compared with the blue graph of finger pulse oximeter device. However, it was worse for EDA data plot comparison.

As Figure 4.9 shows, generally speaking the data graph was not really matching at all. During the skin conductance test, the EDA data increased in the negative high arousal session, remained at a high level in the negative low arousal session and started dropping in the middle of the positive high arousal session. The trend of the changes of the first three sessions were closer with the GSR graph we got from PowerLab 16/35 and Amp set. The rest of the sessions did not match well with GSR data gathered from PowerLab 16/35 and Amp set. We can see clear signs of peaks respond to emotional stimulus from the both graphs of GSR signal however the recorded peaks from E4 did not match peaks with PowerLab through out all the emotional sessions.

It seems E4 applied aggressive filtering as most of the wearable devices do, which decreases the device's sensitivity to small changes by a great margin. Since we did not have access to the raw data, we were not able to tell if E4 gathered useful data before filtering. In addition, the position of the sensors on the E4 for measuring HR and EDA are not the same as the Finger Oximeter or Powerlab used for sensing. The positioning of the sensors has critical impact on the accuracy of the data.

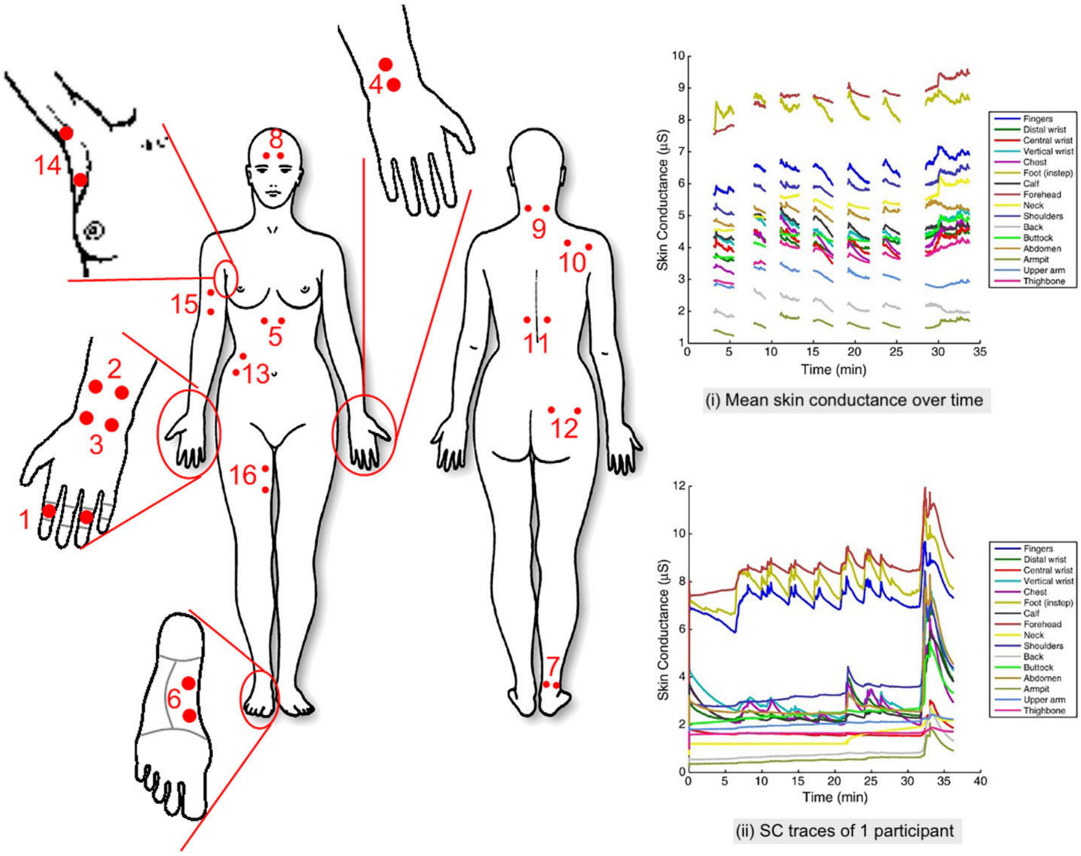

Figure 4.10 16 Skin conductance measurement locations: **1) fingers**, 2) distal wrist, 3) central wrist, 4) vertical wrist, 5) chest, **6) foot (instep)**, 7) calf, 8) **forehead**, 9) neck, **10) shoulders**, 11) back, 12) buttock, 13) abdomen, 14) armpit, 15) upper arm, and **16) thighbone**. Graph (i) and Graph (ii): The vertical axis depicts the skin conductance and the horizontal axis depicts time. [10]

Dooren and his colleagues measured skin conductance level (SCL) and skin conductance responses per minute (SCRs) on 16 different locations on the body (see as Figure 4.10). They also calculated the skin conductance responses amplitudes per minute (S-AMPL) [μS/min] to take into account the strength of skin conductance responses. Among all locations, foot, fingers, forehead and shoulders are most responsive and the most promising for SCL and SCRs recordings [10]. Although forehead was found most high in both SCL, it failed to be in the top 3 regarding S-AMPL. The results showed the foot was most high in S-AMPL, followed by finger and shoulder. They also calculated the correlation assessing similarity with the finger. Among all positions, the foot, thighbone and shoulders were most close to the finger, although thighbone was found to be one of the lowest skin conductance responsiveness.

E4 wristband intended to use the location on the distal and vertical wrist to measure EDA data placing the sensors on the wristband of the device, on the opposite side of the main module. According to the figure 4.10 (i), we can know that the distal and vertical wrist locations are not among the most responsive positions for SCL and SCRs measurement. As for correlations to finger locations, vertical wrist is worse than central wrist and forehead. Besides the location not being ideal for trace SCL and SCRs, the attachment of the electors to the skin are constantly maintained with E4.

In conclusion, wearable devices such as E4 wristband are not ideal wearable device for recording useful data in social interactions for the setting we need in real-world scenario .

4.4. Platform Development

To summarize the above, we were not able to find on the market a hardware-software platform that would fully correspond to the requirements. Also through a series of lab tests with high-end equipment we could not spot any obstacles to implementing such a platform. The sensing locations of interest (EDA, HRV) were sufficiently easy to access from a wearable device and there were several solutions for gathering data from those locations. Given the hardware and software design expertise of our lab we set to build our own platform optimizing it for the following aspects:

- **Cost efficiency.** Price of a device should remain as low as possible to allow the usage of large number of devices. Also it is preferable to keep the computational costs of the software on the low end, in order to be able to use low to mid spec computers to control the recordings.
- **Access to Raw Data.** Platform has to provide raw data sufficiently well synchronized in time.
- **Simultaneous Recording.** The whole recording with multiple subjects has to be controlled from one host, be simple and easy enough to be used by people with no computer science background. Be well synchronized and allow flexible data labeling. In addition to this, the software should be computationally efficient to allow the use of low-mid end computers.
- **Build and placement.** The devices have to be sufficiently well built and not cause concerns of safety and fragility. Wearing the devices should not prohibit normal activities of the subjects.

4.4.1 Hardware

With the above requirements in mind we have chose the form factor of a wristband with sensors placed on the fingers. Finger sensors are connected to the main module with a single thick cable terminated with a 3.5mm phone connector. Thick cable and a single commonly used connector were deemed to be more reliable and less distracting that 3-wire approach that Shimmer employs. To reduce the number of fingers used, the BVP sensor is place on the same finger as one of the EDA electrodes, which allows us to use only 2 fingers. We also use soft elastic sleeves for the sensors, as they are more comfortable than velcro strips that are commonly used.

The smart wristband is equipped with an accelerometer, EDA and BVP sensors (see as figure 4.11). Both sensors are intended to be attached to fingers tightly. EDA is measured using a Wheatstone bridge with 2 electrodes on the fingers measuring the skin conductance. Output of the bridge is fed into a differential ADC, where it is converted into a digital value. BVP is sampled at 50 Hz from a

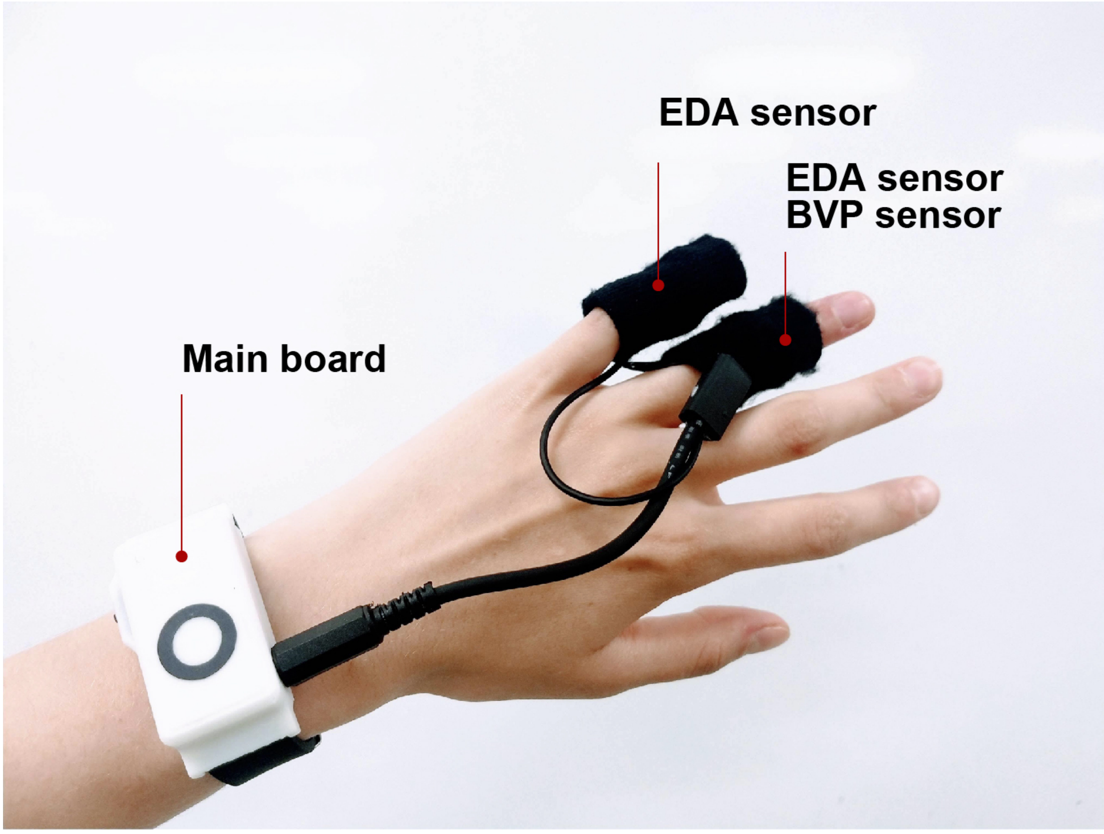

Figure 4.11 Prototype of Smart wristband:The second version of smart wristband with motion sensor, EDA electrodes and BVP sensors on the fingers.

photoplethysmograph. Both EDA and BVP sensors are located on user's fingers. Accelerometer is located on the main board on the wrist. Finger sensors are connected to the main unit with a flexible cable and a 3.5mm 4-pin audio connector. The main unit is based on the ESP32 module and streams the recorded data over WiFi to the server side.

Below is the summary of the hardware optimizations we did in order to match the requirements:

- **Cost efficiency.** The device is build from scratch with all the parts ordered from large suppliers. Except BVP sensors all the circuitry was designed from scratch optimized for small size and simplicity.

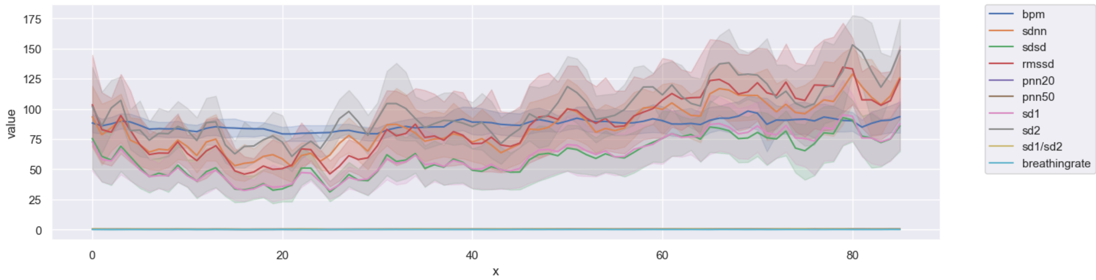

Figure 4.12 Example of HRV feature plot from the data collected through the smart wristband

- **Access to Raw Data.** Since all the sensory circuitry was designed from scratch, the data produced is as raw as it can be. For live visualizations the data is slightly filtered, but the raw data is still available.
- **Simultaneous Recording.** Devices are based on ESP32 controller, allowing them to use IP protocol over WiFi networks. This allows us to simultaneously use multiple devices without worrying about limitations of the Bluetooth stack and without developing our own RF-based protocol. WiFi capabilities allow us to even stream the recorded data nearly in real time.
- **Build and placement.** The devices come in two versions; the first uses a soft wristband and is seemingly more comfortable to wear, the second uses a hard plastic case similar to a wristwatch. In both cases the circuitry is sufficiently hidden from the subjects.

4.4.2 Software

The server side of the platform is run on any network-enabled PC. Software controls each connected device, receives and records the data. Using one central server and a star-like topology innate to WiFi networks, it allows us to start and stop recordings simultaneously and assure the time synchronization and simultaneous labeling of all the recordings. Server side software supports hundreds of devices, which allows us to perform recordings on almost any scale. The possible limitations on the number of the devices come from the arrangement of the WiFi networks and have to be dealt with on the network level. E.g. using several WiFi routers connected to the main hub with a cable link would allow us to nearly double the number of connected devices.

Below is the summary of the software optimizations we did in order to match the requirements:

- **Cost efficiency.** Server side is written in C++ and supports several stages of data buffering. This allows us to avoid computational or memory related bottlenecks. E.g. server does not write to a file every batch of data it receives, but instead buffers it and writes the buffer once it is sufficiently large. This allows to minimize the number of input-output calls to the file system and rely on RAM instead. This allows us to use much slower file storage and much cheaper host computers.

- **Access to Raw Data.** Server records the data as it is coming from the device with no alterations.

- **Simultaneous Recording.** Server heavily relies on the IP protocols to communicate with the devices, which allows to solve the problems posed by the large numbers of the devices through network design. Also software allows labeling and time synchronization of the recorded data. Although the network delays are not addressed, they are too small (in the range of milliseconds) to pose a significant problem for the data analysis.

- **Build and placement.** Server software can be hosted on a laptop. Network capabilities allow placing it away from the experiment site and it is hardly causing any inconvenience to the subjects.

4.4.3 Data Processing Tool

Server software records data in a CSV format with every sample timestamped. This allows using almost any data processing tools of the investigator's preference. However, we provide several approaches in the form of Python scripts that we use for data analysis presented in this work.

The software used for data processing and resulting plotting is based on Python jupyter notebook platform. Since there are plenty of developed Python libraries for signal processing, it allows us to save a lot of time necessary for implementation of such algorithms in our own software. At the data processing stage we can remove the noisy data based on the signal characteristics or the accelerometer data, as well as apply suitable filtering. Pre-processed data is later on used for HRV and EDA feature extraction with variable windowing of the original signal. Calculated features can later be analyzed, rated and categorized based on the hypothesis of the investigator.

4.5. Raw Data Pre-processing and Filtering

Depending on the recording equipment and setup, some data may be affected by noise or be missing. This could present an issue for the analysis. Here we discuss several approaches to mitigate these issues. In general, the noise related issue should be addresses on every stage, starting from the hardware design. On the hardware level it is encouraged to follow the good practices of ground layout, adequate shielding, sufficient bypass capacitance on the power rails, usage of high quality and low noise components, especially ADCs, amplifiers and sensors, and some active or passive filtering circuitry, if necessary.

Unfortunately even following all the good practices and common sense in hardware design, when it comes to physiological sensing on wearable devices, subject's movement can cause the sensors to move and affect the recordings. If the devices are equipped with an accelerometer, it is possible to filter out or remove completely the sections affected by subject's physical movement. Alternatively or in addition to using acceleration data for noise labeling, movement-related noise commonly produces very fast changes in the recordings, that can be easily removed with a smoothing or band-pass filters, isolating the frequency band in which the physiological phenomena of interest are likely to be observed. Indeed, 50-60Hz mains power noise is hardly relevant to the heartbeat, which happens at much lower speeds.

Another common issue with the recorded data is missing data that may occur due to a technical malfunction, lose of network connection, or data being too noisy to be recovered. For large scale simultaneous recordings it is absolutely vital to assure that the data is correctly aligned in time and all the missing segments are padded and filled with zeroes or interpolated between the existing segments. The particularities of this have to be decided based on the analysis method that is being used. E.g. for peak detection needed for EDA or estimation of RR intervals for HRV analysis filling data with zeroes will produce no peaks and thus will have no effect on the results. However, interpolated values may have some effect on the signal's spectrum. If the number of subjects is large it is often easier to omit windows with noisy data than trying to recover something meaningful from it.

4.6. System Testing in the field

In order to improve and perfect the designed platform, We tested the usability and stability of the developed platform in a couple of gathering scenarios such as live performances and concerts.

4.6.1 Dance Performance II

The first version of the smart wristband was tested on a series of dance performance in March 2020. We built over 50 sets of smart wristbands for real time physiological signals tracking. On March 20th and 21st, we collaborated with dancers and designed a dance performance using audience physiological data as part of the performance. In Tokyo, 139 audience members in total attended two days and 3 sessions of performances. We recorded the physiological data of 98 participants over three performances including Blood Volume Pulse, EDA, wrist acceleration and angular velocity.

Participants

Over three sessions, 57, 37, and 45 people attended this performance in total. Most audience members volunteered to participate in our study. We have 98 recordings in total: the first performance 34 (Male= 17; Female= 17), the second performance 31 (Male= 13; Female= 18), the third performance 33 (Male= 19; Female= 14). We received questionnaires from 35 participants in total(Male= 17;

Figure 4.13 Dance Performance: Stage projections of Physiological Data of Audience

Female= 18). By ruling out incomplete or noisy data records, we had 79 (Male= 39; Female= 42) sets of data from the recruited participants for the HRV analysis of this project: the first performance 27 (Male= 12; Female= 15), the second performance 26 (Male= 10; Female= 16), and the third performance 26 (Male= 15; Female= 11).

Figure 4.14 Wristband with electrodermal, heart activity, acceleration and gyro sensors. Wristbands and consent from on the seats.

Procedure

The venue doors were closed 15 minutes before the performance started. All of the wristbands were disinfected with alcohol before each performance and placed on the seat with performance flyers, consent forms and pens (see Figure 4.14).

Before the performance started, the audience was introduced to the concept of this project and the design of the stage elements. The stage elements such as music, on-stage projections and lights were all connected to the audience's physiological data that was streamed live to several computers rendering the audio, visuals or controlling the lights (see the stage visual design in Fig.4.15). The visual projections were mostly made in the form of abstract shapes, where the EDA data was related to the geometry and motion and HRV data was connected to the colors used to render the shape. The connection of the physiological readings and abstract projections, lights and music was explained to the audience using a short pre-recorded data chunk.

Audiences were briefed beforehand on the performance concept, and that they may be asked to wear bio-monitoring sensors. All the audience members were

Figure 4.15 Six sections: (1) Section 1- Suits. (2) Section 2- Business Cards* (3) Section 3- Puppet. (4) Section 4- Romeo. (5) Section 5- Growth. (6) Section 6- Curtain Call. *The photo (2) is a scene while one of the dancer is doing solo with the peak of Bolero instead of the scene of business cards.

briefed about the experimental setup and were asked to read through the consent form on their seats (See Fig.4.14). By signing the form, they agreed to participate in this work and to offer their physiological data. The experimental setup and data collection was conducted according to Ethics Rules and Regulations of Keio University. We had 41 to 50 devices for each performance for data recording.

While helping the audience members to place the wristbands, we briefly introduced how the stage elements would change depending on their physiological responses. 3 minutes before the performance started, we demonstrated how the visuals would change in response to changes in HF/LF ratio and EDA through reference slides as part of the introduction of this performance. The stage elements were designed to be straightforward and engaging for the audience.

Before the performance started, we introduced the concept of this project and the design of the stage elements and physiological sensing to the audience, to explain the concept of the stage elements and physiological sensing. We used pre-recorded data instead of real-time stream, for illustrative purpose. The musical accompaniment consisted of four distinct synthesizer layers. A new layer was added and then modified along with the displayed signal changes. For example, as the HF/LF changed, the specific layer connected to this data had substantial reverb added to it.

Each performance involved seven female dancers and lasted for about one hour. We divided our recordings to six distinguishing sections, each corresponding to one or several major choreographic events, for further data analysis. In the project on the Dance Performance Audience Tracking we are connecting the emotional state of the audience recorded from custom built wrist-worn devices to alter the lightning, visuals and music during the performance itself in real time. Later the recorded data from the participants was analyzed for correlations with the movement of the dancers and dynamics within the audience's behavior.

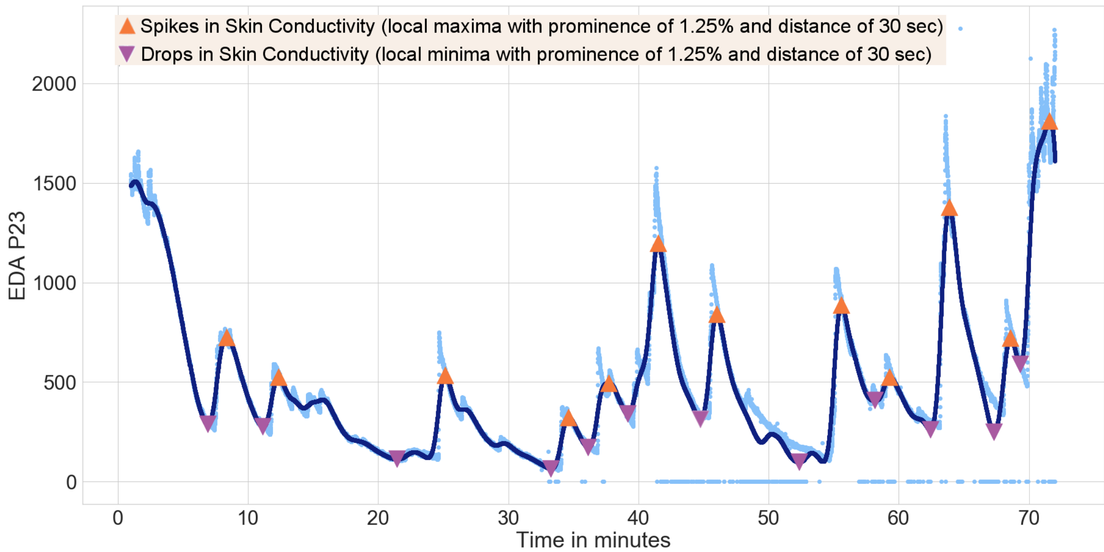

Figure 4.16 EDA data of one of the participants (performance 3, subject 23). Light blue - raw EDA data; Blue - low-pass filtered data; Orange markers - recognized peaks (local maxima); Purple markers - recognized valleys (local minima).

In this project we are tracking and aggregating the physiological signals of the audience to detect emotional response related features from heart rate variability (HRV) and electrodermal activity (EDA). These characteristics are processed and used by the performers to deepen the connection between the audience and the performers, allowing the audience to alter the flow of the performance and thus become a part of it. For this project we are connecting the emotional state of the audience recorded from the smart wristband to alter the lightning, visuals and music during the performance itself in real time. Later the recorded data from the participants analyzed for synchrony across 3 difference performances (see as figure 4.17).

EDA data analysis has demonstrated how such devices can provide insightful data related to the perception of a performance. Each participant's raw EDA data was filtered through a Butterworth low-pass filter from a Python scipy.signal package. Later on we ran spike and drop detection using find_peaks from scipy.signal tuned to detect peaks and valleys with prominence of about 1.25% of the measurement range (0.4095 due to 12-bit ADC) and inter-peak distance of at least 30 seconds. This process detects local extrema with certain characteristics usually associated with emotional responses. Since the venue is located underground and due to the absence of any other audio-visual stimuli and no cell-phone coverage at the venue, we consider all such features to be very likely related to the subject's experience of the performance. An example of one participant's data after processing and local extrema detection is presented in Fig.4.16.

Results

We compared survey results with EDA data we gathered from audience members from 3 dance performances held on March. Among all 37 survey responses, the average satisfaction rating of the performance was 7.63 out of 9. Only 4 participants out of 37 have rated the performance as average or below (5 or 4 out of 9). We found that people who rated the performance very high (8 or 9 out of 9) tend to have a higher number of extrema occurrences during the whole length of the performance (average of 18.13, geometrical mean 17.32) compared to people who gave the performance a rating of 7 or lower (average of 14.17, geometrical mean 13.31). Both ANOVA and T-test comparing the numbers of extrema for subjects who gave a score of 8-9 versus subjects who gave a score of 7 or lower have demonstrated p-values of 0.0075 and 0.0104. Thus, there is a significant difference between the two distributions.

We released the multi-person, multi-modal dataset from in total 98 partici-

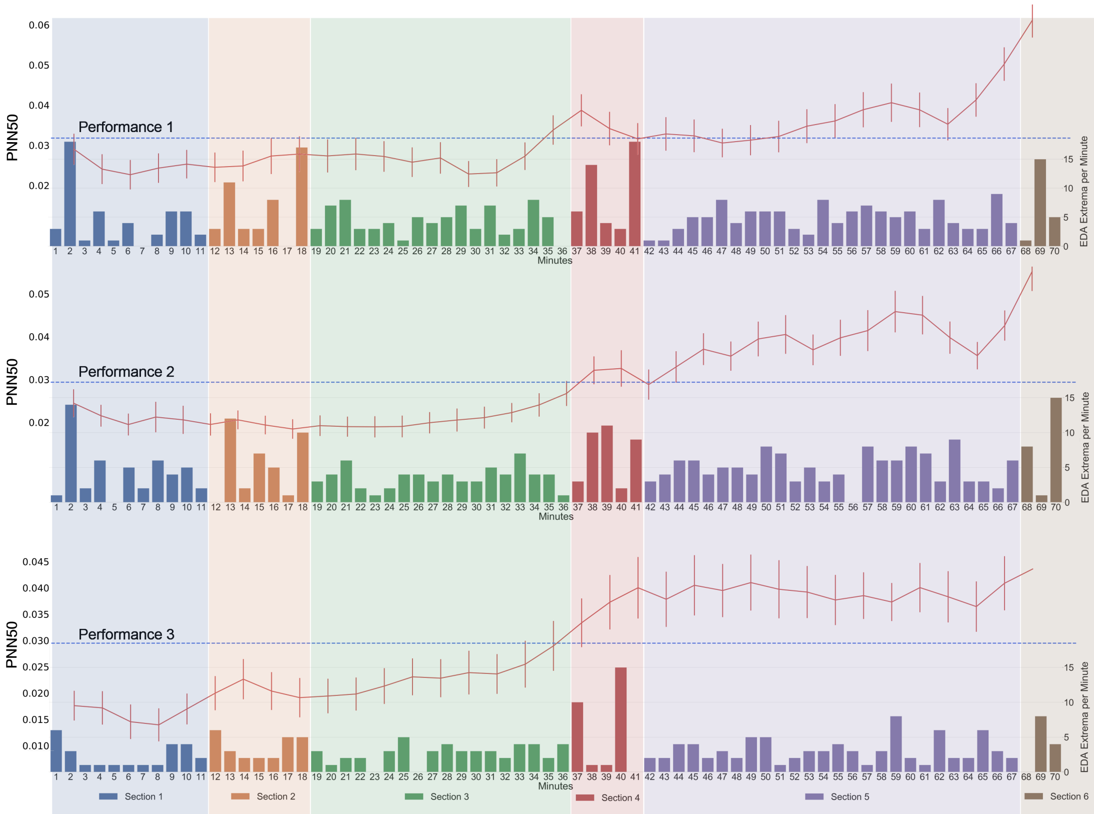

Figure 4.17 Three Performances: EDA and pNN50. Right Y - number of local minima (drops) in the EDA signal. X - time in minutes. Left Y - The change of pNN50. The horizontal blue line represents the average pNN50 for each performance.

pants over three performances . The physiological dataset is 2.5 GB. Together with the sample code this will be made available for researchers under this link: http://bit.ly/audiencebiodata. This project has won the Accenture Interactive Award in the Campus Genius contest 2021 and led to multiple publications at TEI2021 [110] AHs 2021/2022 [111], SIGGRAPH Asia 2022, and WorldHaptics 2021 [112].

4.6.2 Classical Music Concerts

Figure 4.18 Classical Music concert: Frisson Detection and Stimulation Among Audience Group

The first and second versions of the smart wristbands were tested in two classical music concert hold on April and June 2021. In these two events we had 20 and 48 audience members attended the concert. The two conducted large scale musical concerts aimed to test if we can share frisson among audience group with the FrissonWaves system we designed. The system included a thermal neckband as frisson triggers and a physiological sensing wristband as frisson detectors and a trained machine learning algorithm that can process the data gathered from wristband and carry out cold thermal feedback to trigger frisson during the live concert.

CONCERT I

Participants I

In April, the concert had 20 audience members (female=13; male=7) between 20 and 60 years of age (average age=29.5, standard deviation=11.91). To understand live audience's frisson experience, we asked them to fill out a questionnaire before and after the concert, as well as intermission. Based on subjective responses from 20 audience members, 69.72% have experienced frisson and regard it as a positive aspect of an enjoyable live music performance. Wishing to experience more frisson was reported by 83.33%.

Procedure I

The experimental concert included a 40 minutes classical piano program, featuring "Chopin Prelude, Op.28 No.15", "Chopin Ballade, No.1", "Grieg: Piano Concerto in A minor, Op.16 Cadenza & Coda", "Rachmaninoff: Études-Tableaux, Op.39, No.1". These works have a personal attachment to the performer and have a wide range of timbres, textures, dynamics, tempos, and implied emotions which allowed us to observe aesthetic responses of the audiences in many different ways.

Result I As for those who never or rarely feel frisson, all of them reported that they would like to try out a device that can help them experience frisson more frequently. The feedback helped us to understand the audience's expectations and perform the first live tests of the early version of the developed system. In this case the system was not interfering with the performance itself, but simply recording and processing the data.

CONCERT II

This concert was conducted to understand a live audience's frisson experience and find out:

- How does the feeling of frisson occur during a live concert.
- Whether the smart wristband can detect frisson when it happens in live concert.
- How well can we induce frisson experience with our system during a live concert.

Participants II

On the June 25th, 2021 concert at Kawasaki Symphony Hall Assembly Room, 48 spectators were in attendance (female=28; male=19, 1 other or preferred not to say) between 19 and 83 years (average =38.53, sd=15.09). 60% of the audience was familiar with the concept of frisson, 20% were familiar to a certain extent, and the remaining 20% audience members were not familiar with it. On a day to day basis, about half of the audience reported feeling frisson between "sometimes" to "often", while the other half reported feeling frisson rarely to never. However, the numbers of feeling frisson in relation to music are slightly different. 60% of the audience reported to feel frisson in relation to music, and only 40% felt frisson in relation to music rarely to never.

Procedure II Before the concert started, the participants were asked to read and sign the data usage consent form and the photography consent form. Participants were free to leave the concert in case if they do not consent to the terms. 48 audience members were divided into two groups: sharing-frisson and non-sharing frisson groups. The sharing frisson group consisted of 24 audience members wearing our neckbands and wristbands The non-sharing group of 24 audience members wore only the wristbands (Figure. 4.19 shows one example of EDA signal collected by the wristbands during one session of the concert).

This concert included a 50 minute classical piano performance and a 30 minute interactive ambient music performance. The classical piano program consisted of "Beethoven: Sonata No.30 Op.109", "Chopin Nocturne Op.27 No.1", "Chopin Preludes Op.28 No.18 24". This program was selected for wide variety of emotions and reflections on the performer. The interactive ambient music performance, "Reflections on Chopin Prelude Op.28 No.15", consisted of two performers using laptops and synthesizers who utilized the audience's real-time heartbeats and frisson physiological signals using custom-built patchers in Max/MSP. In addition, one violinist and one pianist performed with the electronic artists together with audiences' feedback loop, which transformed the performance from a more structured classical work into a semi-improvisational performance piece directly dependent on the incoming physiological data.

Result II As for the feedback from this concert, the overall enjoyment was rated at 6.2 on a scale of 1 to 7. 62.5% reported that they enjoyed their experience more than other similar events and only 14% answered that they had better experiences at other such concerts. The majority (6 out of 7 people) of them were in the frisson sharing group who were wearing the neckbands. This also matches the participants who found the neckbands to be uncomfortable (7 people in total).

This suggests that a re-design of the neckband may be one of the next steps in order to make it less obtrusive and give a better user experience. Among 24 people in the frisson sharing group, 54% reported that they had more frisson than usually, 25% of them felt less, and the rest found no difference. This result suggests the neckband worked as we expected for triggering frisson in a live concert scenario.

4.6.3 Discussion

The platform worked well in both large scale recordings. The dance performance II in March was the first time we tested the smart wristband V1.0 with a large audience. In Classical music concert II, we tested the smart wristband v2.0.

Platform Modifications

In the new version, we have addressed several issues we encountered while using the original design. To increase the robustness of the device, instead of several thin wires connecting the main board with the sensors we used a thick cable with a 3.5mm phone connector. Thin wires tend to break if pulled too hard, while the thick cable is hardly breakable and will simply disconnect from the device, as the phone connector is not latched or fixed in the socket, thus preventing any physical damage. Another important modification was using a hard plastic enclosure for the device instead of a soft wristband. This way we could conceal all of the inner circuitry and the battery, making the device more fool-proof.

The original version measured the EDA with a fixed-value Wheatstone bridge with an instrumentation amplifier connected to the ESP32's 10 bit ADC. This limits the range at which we can measure the EDA, so very moist or very dry skin was posing a problem for the recording. In addition to this, the built-in ADC is rather noisy. In the new version, we switched to an external differential ADC with programmable gain and re-designed the measurement bridge to allow more flexible measurements. This resulted in much less noise in the data that can be recorded on a wider range of skin types.

One more important change was related to a battery controller allowing the device to charge over USB type C. The early version required a battery swap and did not support charging in any way. Since the plastic enclosure was limiting he physical size of the battery, we managed to improve the battery life by about 50% through some software optimizations. This enabled us to use a smaller battery size while still having a long enough battery run time.

On the software end, we added some more advanced data re-streaming capabilities to the server, as in many applications we needed to process the data and extract certain features in real time. Running the processing scripts on the same machine as the recording seemed rather risky, since data processing may be computationally costly and it is hard to predict whether the server could function properly on an overloaded machine. Thus, the data had to be gathered from all the devices and then transmitted to another machine for processing.

All the changes were rather iterative improvements on the existing design rather than a complete change of direction. This leads us to believe that the project is evolving in the right direction.

Data Interpretation

For the dance performances held in March 2020, we found a clear sync of the EDA data and the PNN50 among the audience members throughout three shows for different scenes of the performance.

There is a significant correlation between satisfaction from the surveys and the EDA peak numbers . This result allows us to note that the EDA data is a suitable tool for estimating the levels of enjoyment or satisfaction of such an event.

Besides, we also found the group EDA data shared the similar patterns over 3 different performances. As shown in figure 4.17, the 3 different batches of audience members shared a very high level of synchrony regarding the same dance scenes. This suggest EDA data can be a good tool to track group feedback over an event as well as to compare the feedback for different audience groups with the same event flow.

As for HRV data, we plotted pNN50 features in figure 4.17. Among 3 performances, the pNN50 features shared a similar pattern of changes before and after section 4 (the romeo scene in figure 4.15 (4)).

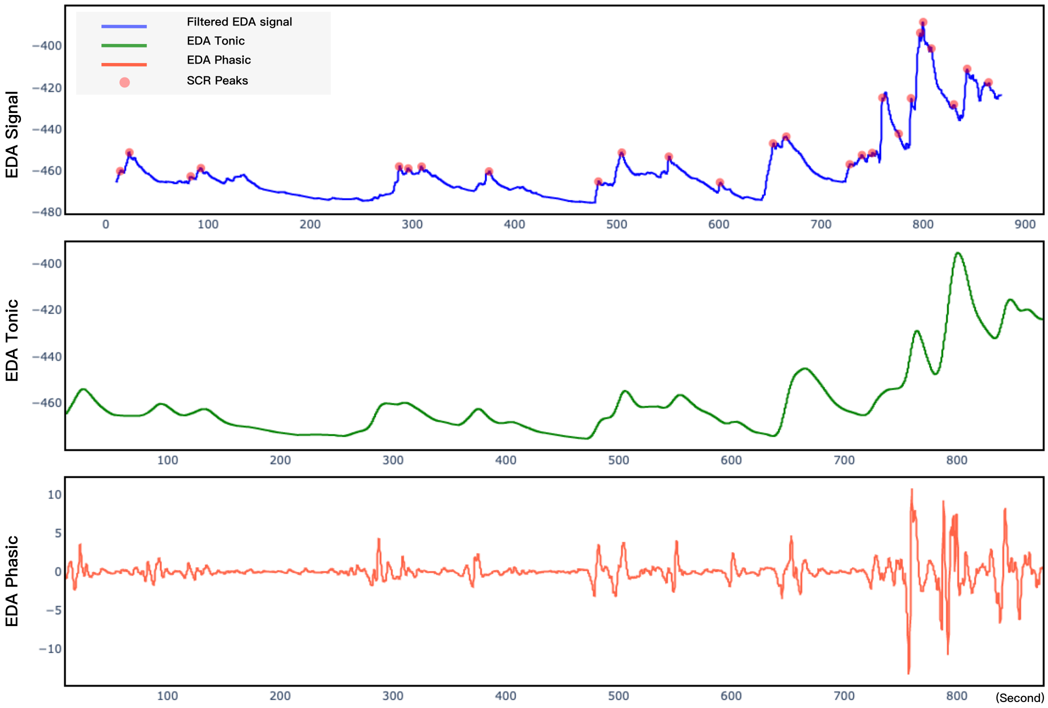

Figure 4.19 EDA signal example collected from one audience member during the concert. The EDA signal plot presents a low-pass filter graph. The orange markers highlight peaks in skin conductance response (SCR Peaks) which is related to sudden change of EDA. The EDA Tonic and the EDA Phasic plots present the components extracted from EDA signal [11] which are key features in the prediction model.

Chapter 5
Field Research

This chapter briefly describes the procedure in which we applied the conceptual approach described in Chapter 3 to field studies of different social events. From section 5.6 to section 5.9 each section presents the participants, study setup and event flow.

5.1. Requirements of the Settings

As we discussed, lab experiments take the human subjects out of their natural environment, yet for relationship and group dynamic studies the response of the subject may not reflect how they behave in their real life. In this project we are interested in gathering valid data of social interactions in a real life scenario where people interact with each other without following instructions of the investigators.

Besides the large scale recordings of performance audiences, there are several requirements for the social scenarios for investigating interpersonal relationships in this project:

- Natural environment of social interactions with self-driven motivation. It is critical for this project that we limit the interference from the investigators and record the subjects' natural behaviors.
- In order to explore interpersonal relations and social interactions, we need at least some scenarios which all the individuals focus on interacting with each other rather than just have a group of people watching a movie together.
- Some degree of deep level communications (e.g., values, worldview, philosophy of life) that people actually can get to know each other deeper rather than a group of people chatting about weather.

- Gatherings involving more than two people, since we are interested in interpersonal relationship among people as well as group dynamics. We need record data from enough people for analysis and association with meaningful interpretation.

- With long enough time duration that we can gather enough data to track the changes of the individuals and group through the event. Most of the social events in real life are not too long in one session considering people generally get tired as the event last. A responsible expectation one session of an event is about less than 4 hours.

- With some basic structure, so that we can arrange and group the recorded data for processing.

When looking for social events matches with these requirements, we found several matching workshops hosted by local organizations and an individual events planner in China. We reached out for collaborations and asked for recording some of their current workshops. We recorded the psychophysiological data of attenders during the events .

5.2. Workshop Collaborators and Coorganizers

Based on the requirement of the settings we reached out to different social event holders, and worked with some of them for this study. In collaboration with a non-profit organization X++ we recorded 4 of their gathering events. X++ intends to connect interesting people from diverse fields together through link-building events including casual chatting, seminars, natural village traveling and so on. They encourage people linked though their events to keep in touch. So far there has been a great number of new projects and events which have been born from a sequence of Friends' Friends workshop X++ arranged. Many of their activities start from a free flowing idea voiced during their events. Many of the attending members also arrange such workshops, events, and theater plays that are open to the general public.

We also collaborated with "Tying Slowly", a bondage performance teaching workshop. The founder of the Tying Slowly workshop describes the concept of the workshop as dance-like body exercise to explore an attentive and compassionate means of non-verbal communication. Although often bondage is associated with sexual activities, they focus on the bondage experience as a means of communication with undivided attention and empathy sharing.

One of the main reason we chose this workshop was, while in an interview, the founder also as the instructor, pointed out that the process of bondage very often reveals how the couple interacts with each other in daily life. That has been said, yet, in their workshop one is not necessary practicing with their partner or even people they know at all. In order to create a safe space, the instructor tends to open the workshop to female students. If a male wish to join, he has to be invited through a female. Some students joined the teaching session with their romantic partners, some come with friends, some just by themselves. The instructor asks all the attenders to pair up with another attender. For attenders who do not bring any partner, they will be asked to find a partner from the workshop scene. It requires to build a high level of trust on the partner for someone to surrender the physical freedom to another individual. Bondage scenarios could create an intimate closeness both physically and mentally. We would like to test how different the experience would be like depending on the nature relationship of the practicing pairs through the gathered data.

5.3. Materials

This section lists and introduces all the questionnaires and materials that were use throughout all the studies for reader's convenience. All the relevant materials are referenced in the corresponding paragraphs.

5.3.1 Measuring Attachment

The Relationship Questionnaire (RQ) The Relationship Questionnaire is a Self-Report Attachment Style rating. It developed from Hazan and Shaver's three categories attachment measure (Secure, Anxious and Avoidant) for adults [123]. This measure consists of four short paragraphs describing the four attachment styles (see Appendix A). After choosing one of the four types, each respondent is asked to make ratings on a 7-point scale of the extant to which they resemble each of the four styles [52].

Revised Adult Attachment Scale (RAAS) [124] is a revised version of the Adult Attachment Scale (AAS) developed in 1990 by Collins and Read [125]. 15 out of 18 items of AAS was borrowed from the Hazen and Shaver's Attachment Style Measure (ASM) [123]; the other 3 items were developed from the earlier work of Ainsworth [126] , Ainsworth et al. [127], Maccoby [128] based on the attachment theory and measurements of infant-care giver relationship [129]. The scale was developed by decomposing the original three prototypical descriptions [123] into a series of 18 items. Theses 18 items contains three subscales measure adult attachment styles named "Close", "Depend" and "Anxiety", each composed of six items. The CLOSE scale measures the extent to which a person is comfortable with closeness and intimacy. The DEPEND scale measures the extent to which a person feels he/she can depend on others to be available when needed. The ANXIETY subscale measures the extent to which a person is worried about being rejected or unloved. In this study, it scored on a 7 point likert-type scale.

Relationship Structures Questionnaire (ECR-RS) The Experiences in Close Relationships—The Relationship Structures (ECR-RS) questionnaire is a self-report instrument designed to assess attachment patterns in a variety of close relationships [130]. The same 9 items are used to assess attachment styles with respect to 4 targets (i.e., mother, father, romantic partner, and best friend). The items were written in a way that allows them to be used for a variety of interpersonal targets (not just romantic relationships) and for a variety of age groups. the research based on the ECR-RS questionnaire also found it indicates that the scales are meaningfully related to various relational outcomes (e.g., relationship satisfaction, likelihood of experiencing a breakup, the perception of emotional expressions), as well as to one another. Method for Assessing Attachment Orientations Across Relationship.

5.4. Participants

Over 5 workshops, we recorded 52 sets of data from the participants (male= 19; female= 33) between 23 and 50 (Mean= 30.4) . Since some of the attendees are organizing events on their own, this led to multiple other opportunities and recordings. We mailed 16 sets of the devices to China for recording, and only for one workshop we did not have enough devices for all the attenders. It took them in average 7.7 minutes to finish the pre-event questionnaires.

5.5. Procedure of the Field Research

We collaborated with several local organizations and individual events planners from three different cities in China for five events, where most of the attenders were asked to wear smart sensing wristbands. With the collected data from the wristbands we were able to track and group behavior patterns while people interact with each other.

In the preparation stage the event host usually will post the event information on their usual platform (e.g., SNS applications, Websites) for recruiting attendees. The application period usually closes once the opening seats are full. If before they cut off the application time, there are still more appliers than the capacity, the host usually will make adjustments and select the attendees accordingly (time availability, commitment levels, field).

The investigator came to the event location around the same time with the event host. With the permission of the host and venue managers, the investigator will set up the cameras before the attenders settling down. In all of the recorded workshops, we ask all the attenders to read through the consent form about the study we are conducting. Once they agreed to volunteer this study, we share a link with a set of online questionnaires to gather demographic information and measurement scales we mentioned in 5.3 depends on which workshop. While the participants are filling the questionnaires, our investigator would help the participants to wear the smart wristbands. Before the event started the investigator will check the wi-fi connection of the wristbands and the computer. The investigator usually will click the start record button at least 5 minutes before the event started.

After the event finished, most of the event hosts will ask for feedback on the event from each attendee through a 5 minute short survey. The hosts shared their survey feedback with us.

Figure 5.1 Field Recording of scenario Friends' Friend I: The Participants are Wearing Smart Wristband v1.0 before the Event

5.6. Field Scenario: Seminar Meeting

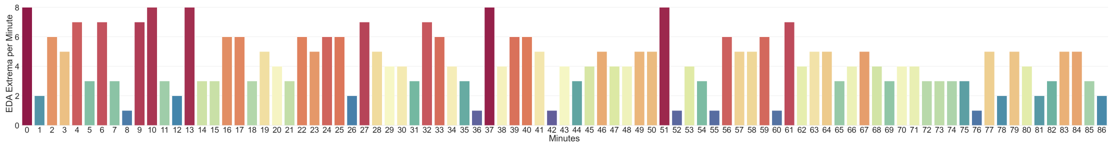

Figure 5.2 Seminar Scenario: EDA Peaks per Minute of all the participants

On November 14th 2020, we collaborated with x++ and arranged a recording for a seminar meeting workshop. This was the first workshop where general public had access to our physiological sensing devices. Once the workshop time was settled, about twenty people showed interest in joining this workshop, due to the participant capacity limitations stemming from the event design only 9 of them were able to attend.

Participants

9 People (Male= 3; Female= 6) aged between 23 and 26 (Mean= 28.7) traveled from Shanghai, Wenzhou, Taizhou to join this workshop at Hangzhou. All of them have higher education with bachelor or master degrees. All the attender were asked to wear the wristbands under the instruction of the host and staff before the event started (see Figure 5.1). 4 of them reported that their hands tend to be colder in general. There is study showed temperature of the skin surface may influence the conductance of the EDA sensors. Thus, we gatherer the self-report information hand temperature of participants.

Event

Most of the attendees joined for the entire day-long event, however we only recorded the seminar session about 86 minutes long. There were one host with other members in Hangzhou and one remote speaker participating from Tokyo over Zoom. In the first half the remote speaker is giving a speech through projection on a white wall while the audience is sitting around a long table. In the second half the host is moderating a discussion on the topics relevant to the speaker's presentation among the audience members.

The event flow resulted to be two main parts: the first half was a online remote speaker sharing from Yokohama while the 10 audience members attended in Hangzhou ; the second half happened when the event host interrupted the zoom speaking because he thought the group were not interested and arranged group discussion instead.

From the data of this workshop we found it interesting to look at the synchrony levels in the physiological data patterns of the members throughout the workshop. There's a higher level of synchrony among best friends (See Figure 5.5).Through out the entire workshop we detected 13 EDA peaks showed in dark red (more than 6 peaks per minute from 9 participants) including 9 group laughs, 4 rounds of group applause after a participant was done with their speech.

5.7. Field Scenario: Sharing with Friends

Friends' Fiend event is introduced by the founder of x++ as a way to get to know inspiring people. People from different fields who may or may not be experts in certain topic gathered together and got to know each other. Usually, these workshops involve 8-10 attenders sitting in a calming or relaxed environment (see as figure 5.3). The main reason they limit the number of attendees for one event is they believe it's a way to make sure everyone has equal chance and about the same time to share.

The event started with a short introductions of the rules. Each attendee will be asked to describe another attendee for 30 seconds to 1 minute. Then the described has 15 to 20 minutes to respond to that description and to share about themselves with keywords they believe that matters to them a lot. The order is randomly selected through the pool of attendees' name cards. As showed in figure 5.4, seat no.4 will have 30 seconds to 1 minute to describe his or her impression of seat No.7. After that No.7 have about 15 to 20 minutes to share before starting describing impressions of No1. and so on. Other members can ask questions, share opinions and join discussions in between. These rules are not strictly applied and it depends on the atmosphere of the discussions judged by the host. However,

Figure 5.3 One of the Friends' Friends events host by x++ in 2021 (photo credit: x++).

most time, the host try to control the flow in situations like when the discussion are around with one individual for too long or lead discussion to individuals who ends sharing too fast.

5.7.1 Friends' Friends I

On November 21st 2020, we recorded a Friends' Friends Workshop in Hangzhou. 14 people applied for the workshop. Due to the time length and the design of the event, the host limited the participants number to under 10.

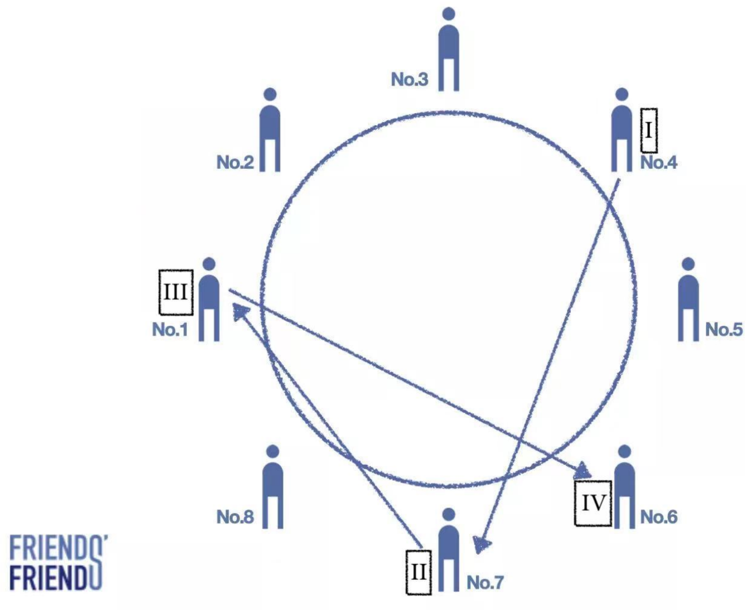

Figure 5.4 The event flow of in general Friends' Friend workshop. Arrows show the order in which participants give a short speech. In this example No.4 starts, then followed by No.7, No.1 and ending with No.6.

Participants

9 of the appliers (Male=2; Female=7) aged between 26 and 38 (Mean= 31.1) traveled from Beijing, Shanghai, Ningbo, Jiaxing to attended this workshop at Hangzhou. All of them have higher education with bachelor, master degrees or under doctor program. 4 of them reported that their hands tend to be colder in general.

Event

The event lasted for about 3 hours. Once all the members arrived, the group started casual chatting till the host started to gather attention from everyone and explained the flow of the Friends' Friend event. With two group discussion sessions in between, all 9 participants shared about themselves for about 150 minutes. After everyone finished discussions, the host asked every attendee to share their feedback regarding their experience of the event. At the end of the event, the founder of x++ showed up and was invited to give a short speech by the host. He introduced the community of x++ as well as the visions of Friends' Friend events and invited people to continue joining the future events.

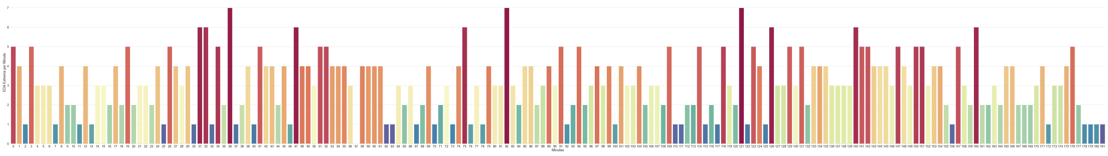

Figure 5.5 Friends' Friends I: EDA Peaks per Minute of all the participants

5.7.2 Friends' Friend II

On December 6th 2020, we conducted a workshop in Wenzhou. Due to the previous events having way more female attendees than male attendees, we discussed with event host and suggesting for a balanced gender ratio. However, we did not interfere with the host's selection process regarding which individuals. About 20 people applied for the event. The host selected 8 of them, in order to allow everyone to have enough time to share their opinions without rush,

Participants

All 8 attendees (Male= 4; Female= 4) aged between 24 to 35 (Mean= 26.8) agreed to be volunteer to this study. Most of them didn't know other people in the event besides the host. All of them are local resident of Wenzhou. All of them have higher education with bachelor or master degrees. 6 of them reported that their hands tend to be colder in general.

Event

The event lasted for about 4 hours. The flow of this event is very close to the friends' friend workshop hold on November 21st at Hangzhou.

5.8. Field Scenario: Board Game

On November 28th 2020, we recorded a board game gathering in Shanghai. Though about 20 people showed their interest to join it, considering the gender ratio and the experience of each attender, 11 of them were selected by the host to join the workshop.

Participants

All the 11 attendees (Male= 5; Female= 6) aged between 23 to 37 (Mean= 29.2) agreed to be volunteer to this study. All of them are resident of shanghai. All of them have higher education with bachelor or master degrees. 6 of them reported that their hands tend to be colder in general.

Event

The data recording of the workshop is 167 minutes long. The host of this event brought a board game called "10 thousand times dating", in which the players will experience confessing to other players or being confessed to. This board game was designed to stimulate romantic atmosphere among players. However, the host of this workshop found the participants paid more attention to winning the game rather than romantic interactions with other players.

We sought to test the host's judgment. On one hand, it will not be a surprise if the host was right; on the other hand, if some of the players are experiencing differently from what host described, it would be interesting to analyze what is different, how different and why different. However, if the host is right there may be a lot potential explanations. For one thing, people nowadays may be more interested in competing with others and being a winner than developing romantic relationships.

5.9. Field Scenario: Bondage Performance

On November 29th 2021 in Shanghai, we recorded a bondage workshop. 22 people attended in this workshop resulted in 7 romantic partners pairs, 1 female friends pair, 1 stranger pair and 1 individual from a female friends pair. We only had 15 working devices available for this workshop.

Participants

15 of the attenders (Female= 10; Male= 5) aged between 25 to 50 (Mean= 33.9) volunteered to wear the wristbands through out the event. We recorded 7 paired up dataset and one single recording from our wristband v1.0.

Event

The workshop lasted for 155 minutes. In this workshop there were mainly three sessions. The first session was just observing a complete bondage performance from the instructor and his model (also his romantic partner) followed with a body exercises preparation. Next, it started with the teaching session of hugs. Students were taught how to mimic full embraces to their paired partner with ropes. The last session was a more advance trying session where the instructors demonstrated how to practice a back hug with ropes. Meanwhile he encouraged students to explore more with their pairs.

We intended to have the attendees to feel as secure and relaxed as possible. Therefore we did not ask for the consent nor the right of taking or publishing the photographs of the attenders face.

Figure 5.6 Field Recording of scenario Friends' Friend II : at the end of the event the attenders took a picture of their hands wearing the psychophysiological sensing wristbands.

Figure 5.7 One of the bondage workshop hosted by Gandalf: The instructor and his partner model presenting a bondage performance (Photo credit: Gandalf)

Chapter 6
Results and Discussion

This chapter presents the evaluations of the data results we gathered from the field studies described in Chapter 5 in which we implemented the proposed approach outlined in Chapter 3. The sections start with how we looking into the data patterns that reflect behavior features and social reality for possible predictions, followed by the discussion of the recorded dataset, analysis and evaluation method. We have recorded 1.1 GB of physiological data of 52 participants from 5 workshops, and 26.2 GB of video recording.

6.1. Result of Questionnaire

The result we received from the relationship questionnaire (RQ) is rather interesting. The participants were asked to choose one style out of four styles that fit them best in descriptions followed with four ratings assessments on a 1-7 points scale of the degree to which each of the four styles fit them (show in A).

In the beginning question of RQ, 21 participants (40.38%) chose secure style (A), 13 of them (25%) chose dismissing style (B), 9 of them (17.31%) chose preoccupied style (C), and 9 of them (17.31%) chose fearful style (D).

Nevertheless, in the rating questions, 15 participants (28.85%) rated highest in secure style (A), 4 people rated equal high of A and another style, 1 rated equal high for A, B, D and, 2 rated equal high for all four styles. Moreover, not all of them are from those rated A in the beginning. The same can be found in dismissing style (B) rating. 16 Participants (30.77%) rated highest in dismissing style (B), 5 people rated equal high for B and another style. 4 Participants (7.69%) rated highest in preoccupied style (c). all of them also chose C in the beginning. 3 participants rated equally high in C and D. 2 of them chose c and 1 of them chose D in the beginning. 4 participants (7.69%) rated highest in fearful style (D). 3 rated equally high in B and D.

Attachment Styles	1st Label	Exact Match		Mismatch		Multi-Match	
Secure	21	13	61.90%	4	19.05%	4	19.05%
Dismissing	13	10	76.92%	1	7.69%	2	15.38%
Preoccupied	9	4	44.44%	1	11.11%	4	44.44%
Fearful	9	3	33.33%	3	33.33%	3	33.33%
Total	n=52	30	57.69%	9	17.31%	13	25%

Table 6.1 RQ Attachment styles: self-report Labeling and Rating Matching Results

As showed in table 6.1, there are only 30 out of 52 (57.69%) participants completely matched their choices of attachment styles in both sections of the RQ (Exact Match). 13 participants (25%) matched more than 1 type in their rating but all of them have included the beginning one style choice (Multi-Match). 9 out of 52 (17.31%) mismatched their choices in the two section of the same RQ (Mismatch). 6 of them (66.67%) were highest on dismissing style (B), but they chose other style in the beginning RQ questions, 2 of them were highest on secure (A), and 1 of them were highest on Fearful (D). Among the misjudged dismissing style (B), half of them labeled themselves as secure style (A) in the beginning question, the other half labeled themselves as fearful style (D).

The average time they spent answering the questionnaire were 7.7 minutes. Most of participants finished it within 11 minutes, all of them finished within 25 minutes. Only 3 participants took more than 15 minutes. The RQ only contains 5 questions out of 33 questions. Since the questionnaire were also filled in a field study environment, we did not completely control the attentions and limit the tasks of the participants in the middle of their answering process. Some of them may answer a emergency call while the time counting of questionnaire ongoing.

However, it is clear that within less than 10 minutes 17.31% of the participants' judgements are not consist in the questionnaire answers in the first multiple choice question and the following 4 rating questions.

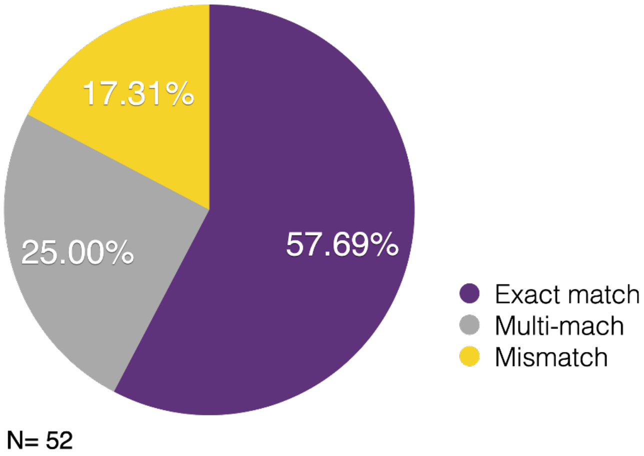

Figure 6.1 Self-report Inconsistency of Relationship Questionnaire (RQ) in two ways of asking the same questions

6.2. Psychophysiological Data Interpretation

In order to avoid the expectation bias from the researchers, we applied double blind methods in analyzing the psychophysiological data. We plot the data set without knowing the particular events and participants in the workshops. If we find a distinguishable pattern in the data, we check if it is reflected in the video and the data we gathered through questionnaires, feedback surveys, and interviews. Thus instead of looking for reflections of known facts and events in the data, we explain the discovered patterns and anomalies with what we know about the workshop from other sources. This approach is inherently less biased.

Establishing the correlations between features of data from traditional methods and patterns of physiological sensing data, can make psychophysiological sensing method a viable alternative when experimenters have difficulty to use traditional measuring methods.

By combining the physiological data with traditional methods, we extract additional information that would not be available through the traditional methods alone. For instance, in almost all the events we recorded we gathered the group feedback on how much did the subjects enjoy the event afterwards. However, it was not detailed enough to tell which part of the activity they liked more than the rest. The information regarding that is rather blurry if there is any at all. It is a natural phenomenon for people to lose track of their feelings and emotions and not be able to recreate accurate memories. Psychologists have offered more than enough evidence to show how bad we can be at recalling facts and how misleading our memory can be [131–133], even vivid memories can be false or lack of accuracy [134, 135]. The video recording can help us capture what happened but not exactly how everyone felt during the event. With the psychophysiological data recordings, we can tell in which exact moment people are more excited compared to other times. Based on the time information and matching it with videotape we can find out information on what are were the topics and whose speech the group liked more.

6.2.1 Features of EDA

Within this work we used several approaches to interpret the EDA recordings. Brief description of the method and its meaning and potential interpretation are discussed below.

Peak Detection Based Approach

Initial and most straightforward approach is based on extracting spikes in the EDA recording. Depending on the sensor placement and the recording equipment spike parameters may have to be adjusted, but in general, the approach requires labeling all the spikes with certain prominence within a few second window. The number of these spikes per unit of time (e.g. minute, for easy interpretation) can be a rough estimate of phasic EDA. The larger the number is, the more aroused the subject is. Unfortunately due to large differences between subjects it does not seem to be possible to establish a strict relationship based on the number of spikes alone. But given a long enough recording it is possible to establish a within-subject baseline and see whether the subject is more aroused or less aroused than average for each particular minute (or any other bin size).

Additionally this approach makes it very easy to average the results over a large number of subjects, as we can simply average out the number of spikes per any given window size. This allows to see the most exciting or stressful sections of

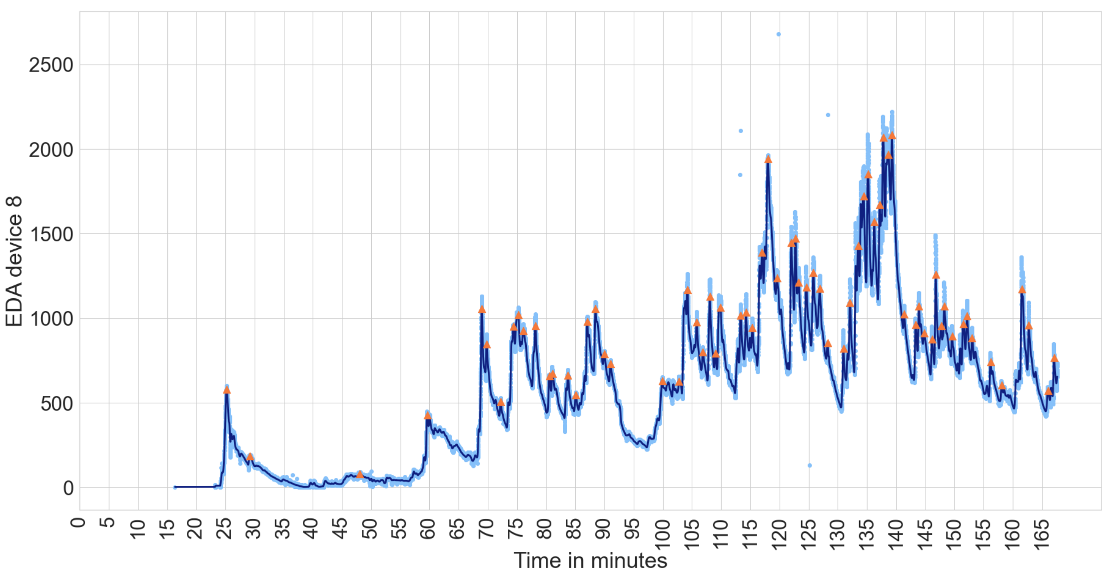

Figure 6.2 EDA peaks detection features from one participant attended Board Game event.

any given group activity. Main advantage of this method is it's computational efficiency, as peak detection does not require filtering or Fourier transforms. On the downside it may be susceptible to high frequency noise, producing spike-like forms in the data. This downside however can be addressed by more careful spike detection, using more than just prominence and window size, but also the shape of the peak. As EDA peaks have a specific shape that is rather unlikely to occur if the measurement is affected by noise of some nature.

Spectral Analysis

As the name suggests, is based on frequency domain. Phasic component of the electrodermal activity happens much faster than the slow tonic changes, and it is the phasic component that we are more interested in, when talking about dynamic experimental conditions, such as social environments. It is reasonable to attempt isolating the fast changes from even faster noise-induced signal components and the slow tonic changes using Fourier transforms. If the signal is broken down into frequency bands with their respective powers, we can easily isolate only the bands related to phasic changes and see how much activity within only this band happens at any moment in time using a spectrogram. In this work we used SciPy Python package for fast Fourier transofrms and spectral analysis [136]. Using this library we calculated the absolute magnitude of the Short Time Fourier Transform over the power spectrum.

As a result we get information very similar to peak detection approach, but much less influenced by noise and other artifacts. However this method is more computationally costly and aggregating data among multiple subjects is not as straightforward. But on the other hand it captures more details and contains more information.

Synchrony: Signal Coherence

Since the EDA recordings can be interpreted as signals, it is possible to estimate their coherence, as a metric of their similarity. Signal coherence expresses how much power of signal A is transferred into the signal B. Although obviously there is no EDA transfer happening between subjects, if their reactions to the same event are similar, we will see an increased coherence in the frequency bands related to phasic changes. In other words it can serve as a good metric of syncrony between a pair of subjects and show whether they react similarly or not. In this work we used SciPy Python package Coherence, the normalized cross spectral density [136].

6.2.2 Features of HRV

Heart rate variability (HRV) describes the tiny differences in time between each heartbeat. It consists of changes in the time intervals between consecutive heartbeats called interbeat intervals (IBIs) [137]. "R-R intervals" describes interbeat intervals between all successive heartbeats. N-N intervals refer to the intervals between normal R-peaks [138]. From the data gathered through BVP sensors, we applied time-based features of HRV such as BPM, SDNN, RMSSD and pNN50, Frequency based features HF, LF, Non-linear features such as: SD1,SD2, SD1/SD2 and Breathing related features such as RSA and breathing rate (see Figure 6.3 and 6.4).

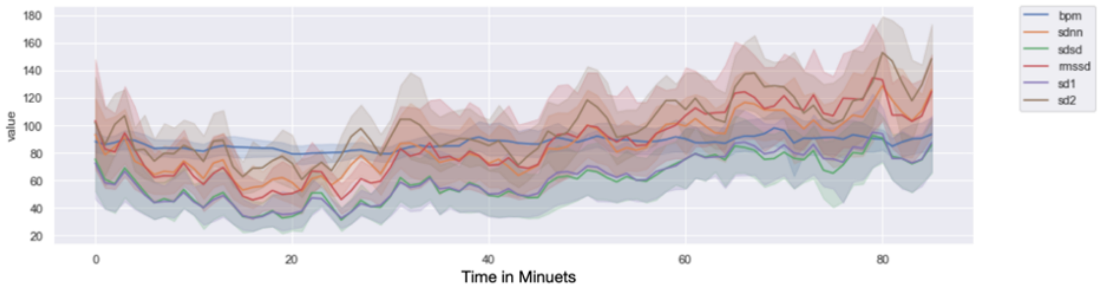

Figure 6.3 HRV features plot from participants attended seminar event: Navy Line– bpm, Orange– SDNN, Red–RMSSD

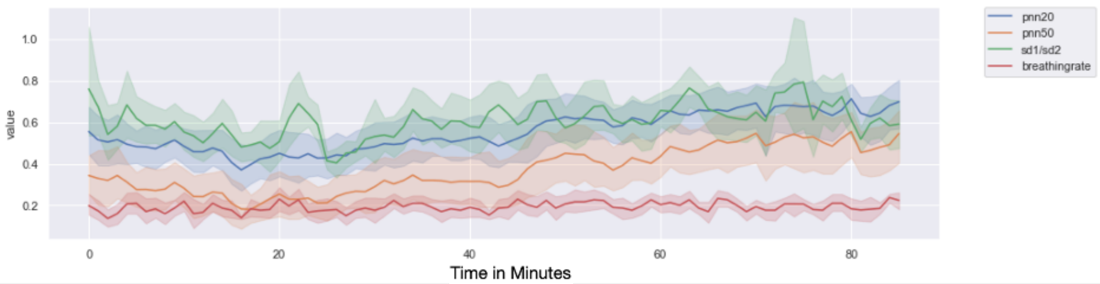

Figure 6.4 HRV features plot from participants attended seminar event:Green Line– SD1/SD2, orange– pNN50, Red– Breathing rate

The HRV metrics are calculated for a time window, for the present work it is usually 2 minutes or longer, with 50% overlap. Meaning that each successive window starts in the middle of the previous window. This greatly reduces the amount of data and gives us the HRV values updated every minute (due to 50% overlap). Since the recordings are syncronized, data for each minute can easily be aggregated and averaged among all the subjects. It allows us to see a rather detailed image of whether SNS or PSNS is more dominant and whether the subjects are calm and at rest or aroused and stressed.

We have discussed most of these features at 2.3.2, besides breathing rate. Breathing rate can be extracted from HRV data based on respiratory sinus arrhythmia, or the fact that during inspiration the heart beats faster and slower during expiration. Thus, based on heart rate we can have a metric of how fast the participants breath during the event. Figure 6.18, is a single feature plot of group breathing rate. The higher Y value is, the faster breathing rate is. From this graph we can see that during certain moments the entire group breaths faster (light blue part) and sometime the entire group breathing slower (grass green and purple part).

Synchrony: Dynamic Time Wrapping (DTW)

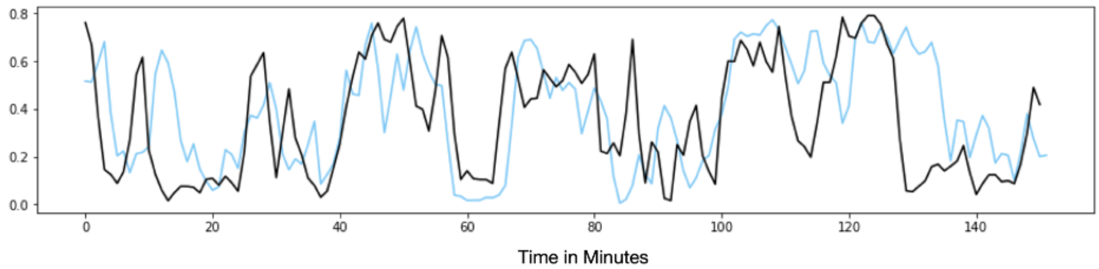

Figure 6.5 pNN50 of the romantic couple pair D7 (blue) and D8 (Black) during entire bondage Workshop

We used Dynamic Time Wrapping for comparing the HRV features. With that we could know how close two participants are similar to each other in there HRV physiological signals. DTW was calculated using a Python port of the R DTW library [139]. We use the following parameters for the best performance: Window type: slanted band, Window size: 5 (min). Step pattern: symmetric2. Window size of 5 samples (one sample per minute) allows us to see the synchrony between participants's signals with up to 5 minute delay.

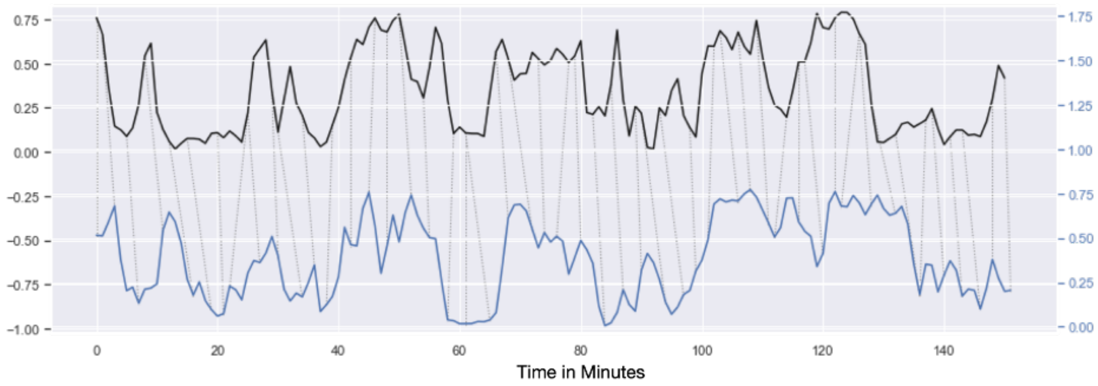

Figure 6.6 the pNN50 Dynamic Time Wrapping of the romantic couple pair D7 (blue, Y value in the right) and D8 (black, Y value in the left) during entire bondage Workshop. The DTW of this graph is calculated in the window of 4 mins. The pNN50 distance between D7 and D8 is 23.521

6.3. Excitement Degrees

6.3.1 Group

In the seminar field study, according to the video, the first half was about 40 minutes long and the host thought the audience members were losing interest, so he suggested to proceed to the group discussion part. However, the data we gathered did not align with the host's judgment. According to the EDA data we gathered from all members, there were more peaks in the first half during the online seminar and a very clear drop at 40 minute while the zoom was cut off and remained low amount of peaks when the seminar shifted to group discussions (see as figure 6.7). The distinguished drop between the first half and second half suggests the group found the discussion part less interesting than the first remote speech session.

One of the traditional method in psychological and social research is having the participants activities recorded to later be reviewed, categorized and/or rated by invited experts or trained volunteers. It is difficult to deny that this method and self-report feedback not only can be very subjective but it can also be deceptive,

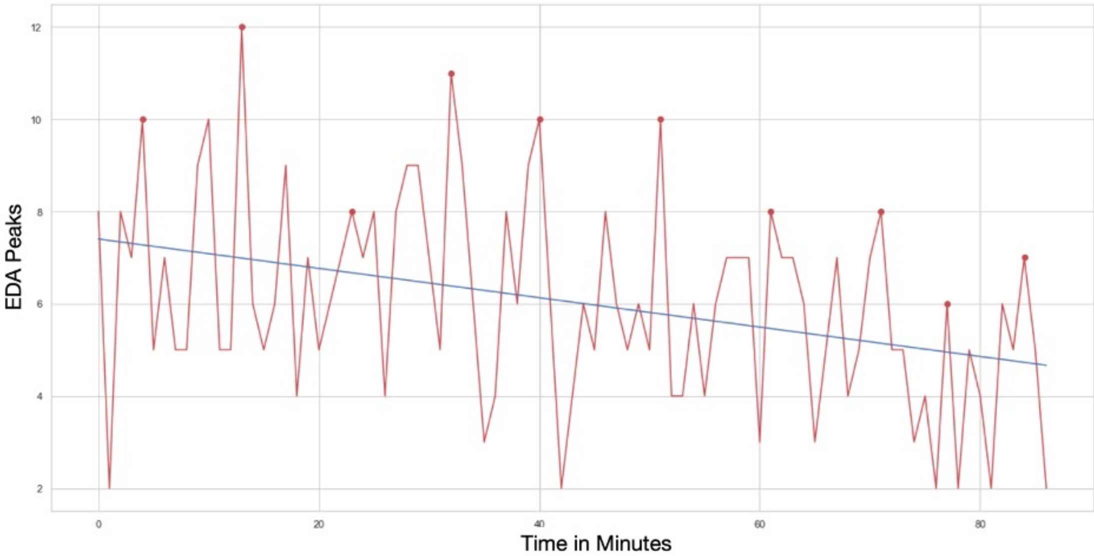

Figure 6.7 Seminar Meeting: EDA Peaks through out the afternoon Session. the first 40 minutes was remote lecture through video call while the rest was discussion among attenders.

as only visual and audio modalities recorded from a fixed angle are available.

When it comes to social and psychological studies, one of the first lessons to learn is what appeared to be true may simply not be the case. This workshop reveals how a very experienced host may still misjudge the event atmosphere. We consider conducting a follow-up study having a third party reviewers (besides the audience and the speaker) to judge based on the video recording of the workshop, and see if they will or will not make the same judgement with the host. We wonder if they may come to these understandings:

1. What it appears to be may not represent what actually going on in this particular workshop
2. Having a third party to read the behaviors and atmospheres of the event environment can be as deceptive
3. Using physiological sensing can cover this blind spot for study used to only have taping-reviewing way for evaluation.

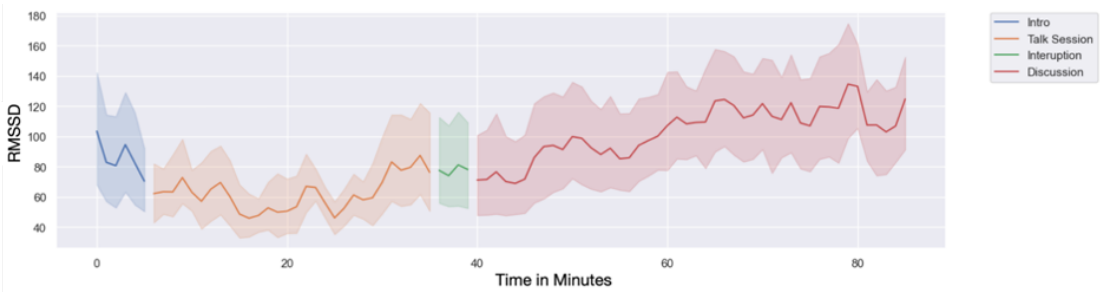

Figure 6.8 Seminar Meeting: HRV RMSSD Feature. Different color areas represent for different sessions during the event.

6.4. Synchrony of Interpersonal Relationships

In the bondage workshop We found the synchrony of the heart rate variability is the strongest among romantic couples followed by friends, and the weakest for strangers. In figure 6.9, we plot three graphs of RMSSD features from three different pairs during the bondage workshop. Graph (a) was generated from a pair of female attenders who didn't know each other before this workshop, graph (b) was generated from a pair of female friends attenders while graph (c) was generated from a female and male romantic couple attenders. From these 3 graphs we can tell the synchrony levels of these 3 pairs based on how similar the RMSSD curves of the pairs are through out time. As showed in figure 6.9 we found paired participants during the bondage are relatively small in pair (a) where the two participants are strangers, then get more prominent in pair (b) where the two participants are friends, and very obvious in the pair (c) where the two participants were a couple. This initial finding suggests that it is possible that the closer level of intimacy the pair has, the higher is more synchronized their responses are, reflecting that they may share the same emotions during the bondage workshop

From figure 6.10, we can see, comparing to the pair 9 and 10 , the coherence levels of the EDA signal from the pair 12 and 13 were higher on almost all frequencies. The EDA data analysis is consist with HRV data analysis. As for couples has lower synchrony patterns on HRV features such as RMSSD also showed low in EDA frequency coherence plotting. Even though the coherence of 9 and 10

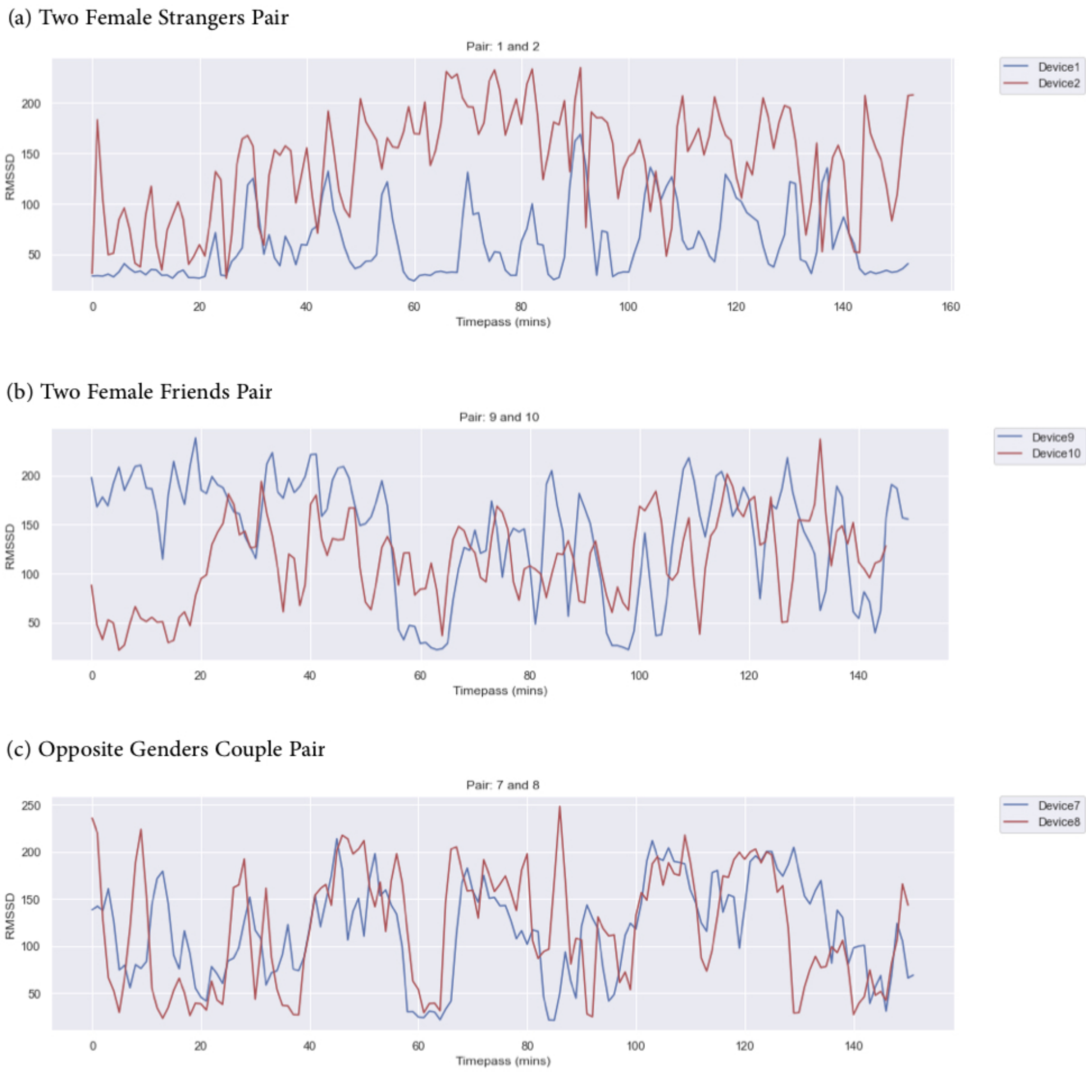

Figure 6.9 Bondage Workshop: The RMSSD of 3 pairs of participants: (a) Two female strangers, (b) Two female friends, (c) an opposite genders couple

Table 6.2 Dynamic Time Wrapping Distances of RMSSD in Three Practices during the Bondage Workshop

Pairs \Sessions		Whole	Practice 1	Practice 2	Practice 3
Non	1&2	14809.015	2592.965	1194.037	1677.943
Romantic	9&10	9698.374	456.477	716.754	666.283
	5&6	7954.753	841.219	807.193	315.823
Romantic	7&8	5277.274	273.042	376.754	561.421
Couples	12&13	7426.770	498.902	861.644	748.468
	15&17	6493.361	564.144	678.994	730.329

The p-value is .021551. The result is significant at p < .05.
Practice 1: 61-75 mins, *Practice 2*: 89-102 mins, *Practice 3* : 137-152 mins.

are lower than the 12 and 13 pair, it's still relatively high knowing the random coherence shall be near to 0. In other words, the result of their coherence level reflects that the participants wearing devices 9 and 10 were attending the same event as a pair.

We ran Dynamic Time wrapping distance analysis when 6 pairs were practicing with each other in three practice sessions of the workshop while 61-75 minutes, 89-102 minutes, and 137-156 minutes of the recordings (see as Table 6.2). We filtered out the pair of the instructor and his model (D14 and D16)since their demonstration time was different from the rest of the group. Besides, considering the nature of their practice scenario were very different from the rest of the pairs. The result of ANOVA test showed rmssd The f-ratio value is 6.48521. The p-value is .021551. The result is significant at p < .05.

It is a very small sample size to conclude anything generalizable. Yet combining all the above findings together, the initial result of physiological data suggests that it is possible that the closer level of intimacy the pair has, the higher is more synchronized their responses are, reflecting that they may share the same emotions during the bondage workshop.

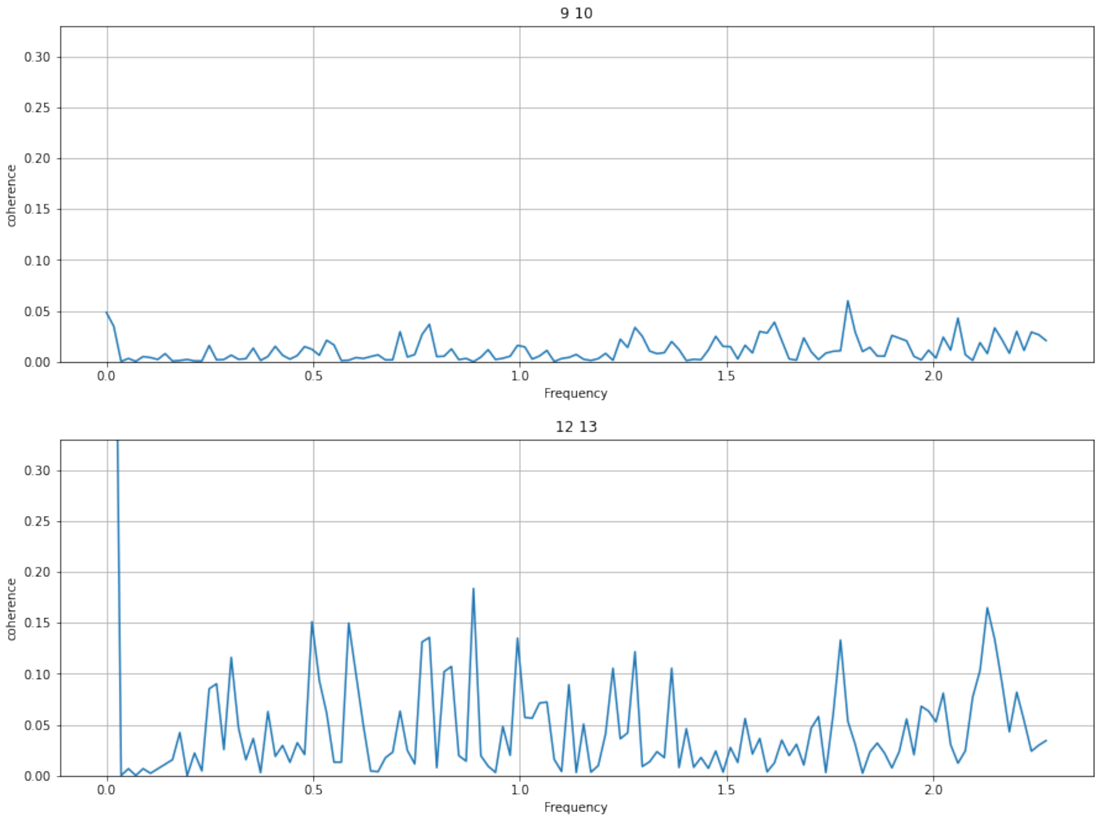

Figure 6.10 Bondage Workshop: Coherence of 3 Pairs Based on Frequency of EDA signals

Leading and Following

Figure 6.11 is a graph of pNN50 of pair 14 and 16 under dynamic time warping analysis. This graph can not only show how emotions similar to each other between the pairs but also who's emotion is leading and who's following.

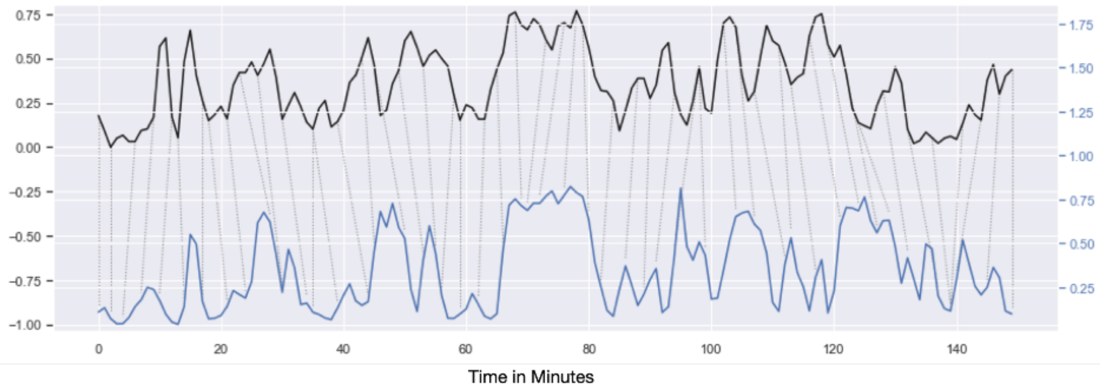

Figure 6.11 Dynamic Time Warping of pNN50 Features of the Pair D14 (Blue) and D16 (Black) through out the Entire Bondage Workshop

6.5. Dynamics among the Group

Together with the traditional psychometrics we found interesting features of data patterns.

6.5.1 Individual Differences

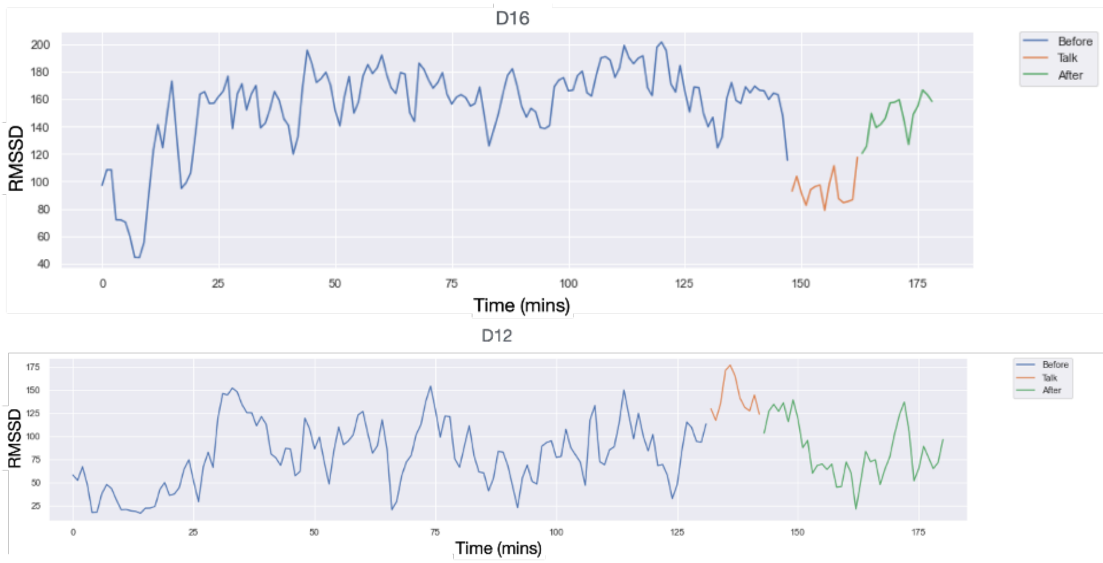

Figure 6.12 Friends' Friends Event 1: rmssd Feature of Participant who wore Device 16 and Device 12. Orange area is while the Participant is sharing, Blue is before sharing and Green is after.

Though all the member attended the same event we can tell through physiological data feedback the behavior pattern or their cognitive process regarding similar situation are different.

From Graph 6.12, we can see participant wearing D16 was very relax during the entire event beside the time when she won the lottery and became the person to share. After her sharing her rmssd value went back to high level meaning she felt less intense and relax again. The participant wearing D12 had opposite responds during the talking session. She got more chilling comparing to before and after her

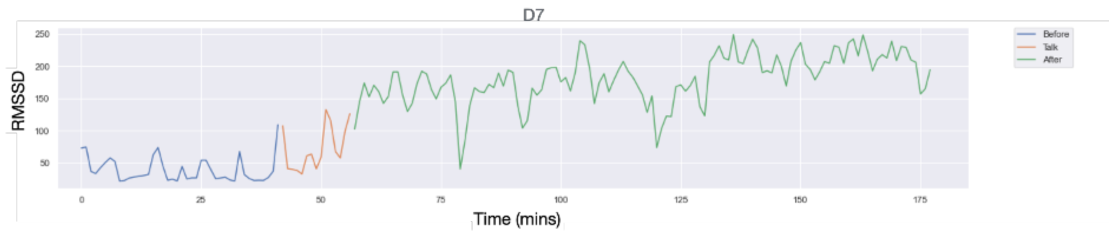

Figure 6.13 Friends' Friends Event 1: HRV rmssd Feature of Participant who wore Device 7. Orange area is while the Participant is sharing, Blue is before sharing and Green is after.

talking session. The participant wore D7 had also some difference and similarity experience with D16. From graph 6.13, we can see, as D16 this participant also became more relax after her sharing session, however she was remain rather intense or less relaxing most of the time before her sharing time. There is a very clear drop when she was announce to be the next person to share and the previous attender was describing of first impression of her. However, her own talking time was only the first 2 minutes. The host got involved and encouraged the group to discuss over her topic, hence more than 70% of her session wasn't really her own talking time. We can see she got more relaxed as it went till the ending where she would be asked to lottery the next member and describe the first impression.

In Friends' Friend II, event, we also found some interesting dynamic from the HRV data. As showed in figure 6.14, the graph of the HRV feature showed a mirrored pattern between data gathered from device 8 and 7. Over certain topics one individual became very intense while another individual became very chill at the same time.

In-group? Out-group?

In the board game event, we processed the HRV data with comparison matrix of all attenders to see how similar the data sets is to each other. The researcher was blinded from the event as well as the information of the participants. As we can see from Figure 6.15 sdnn left graph, there is a blue square block in the

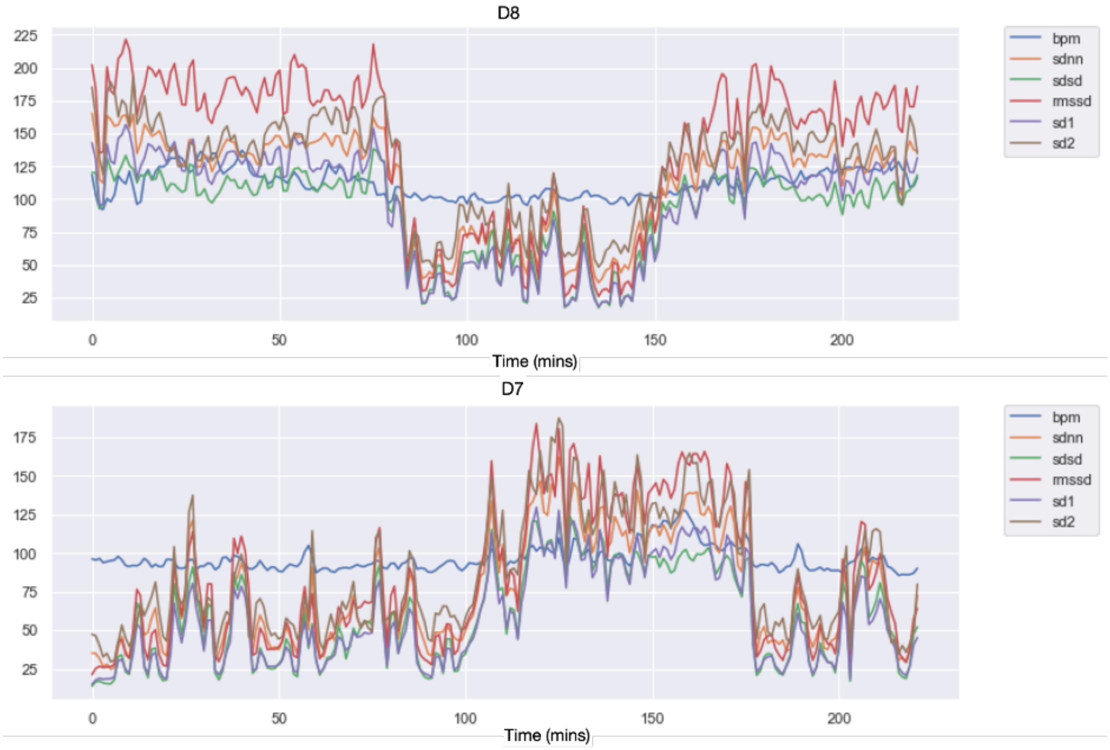

Figure 6.14 Friends' Friend II: HRV features of D8 and D7 during the event.

upper-left and a blue stripe between D10 and D1, D2, D4 and D5. This indicate there a very strong synchrony levels among 6 individuals who wears device 1, 2, 4, 5 and 10 during the event. In the figure 6.15, we can also see D9 was in a very different from all the members itself from both SDNN and RMSSD. When we checked the questionnaires, what device 1, 2, 4, 5 and 10 shared in common were they all know the event arrangers who were in the event that day. Meanwhile the individual wearing device 9 was the only member didn't know anyone before the event and she was late for the event. There maybe other explanations of the different synchrony levels, nevertheless, we can find the synchrony levels through processing EDA data which also reflect the group dynamic of the participants are part of the contribution of this work.

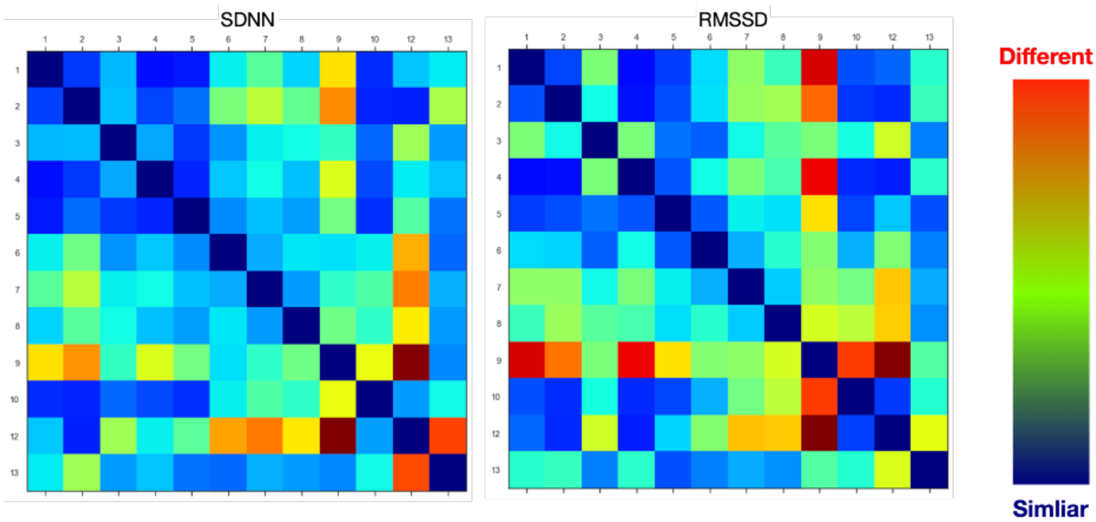

Figure 6.15 Board Game event: the synchrony comparison between all individuals of SDNN and RMSSD HRV features. The closer to darkest blue in the graph the higher level of synchrony it is between the two data sets, while the closer to red the higher level of differences between them, e.g. the data of 1 and 1 are identical–100% synchronized, hence the color block of 1 and 1 is the dark blue.

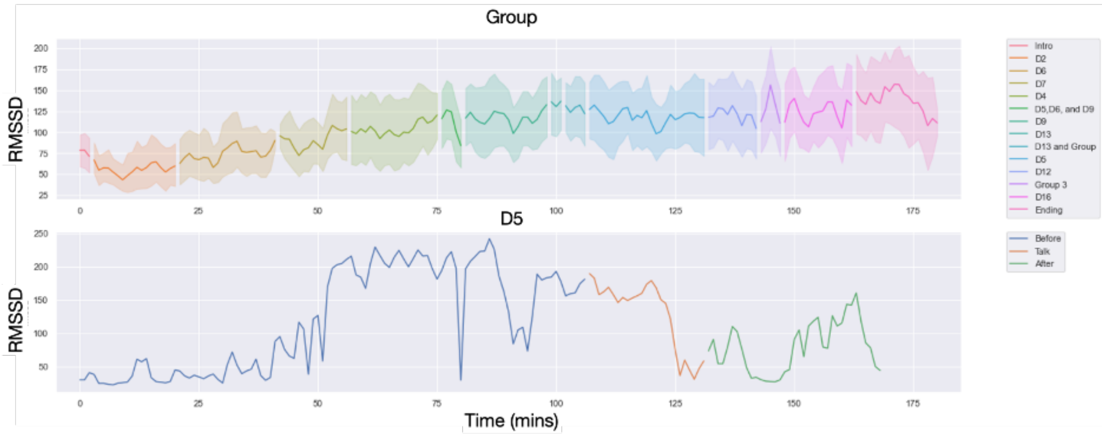

Figure 6.16 Friends Friends I: RMSSD features of the group (up) and subject wearing device 5 (down) in this workshop. X axis in both graphs present for the timeline of the event. The upper graph is a RMSSD feature of HRV. Different colors indicate different parts of sharing and discussing session during the event. The different colors of graph in the bottom indicate different RMSSD value during before , during and after talk time of D5

Popular? Unpopular?

In the Friends' Friend I event, after researcher processed the HRV data. We found a distinguish change of HRV feature in a particular time. Figure 6.16, is a graph of HRV RMSSD feature of the entire group through the Friends' Friend I event. As we discussed before the raising value of Y indicate the group getting more relaxed while the drop meaning more stressful. From the figure6.16 RMSSD, we can see a clear drop once the participant wore device 5) started sharing, and remain low till the end of the session. Once the participant wore 12 started, there is a very big raise. These changes indicate the group got more stressful when participant wore device 5 was sharing and went back to calm and relax once he finished. However, during the same timeline the individual HRV data collected from devices 5, were very different from the group data. The figure6.16 bottom group is all the HRV features of participants wore device 5. During the 5 sharing session (blue area in the top, orange line in the bottom) we can see the Y value went up and remained

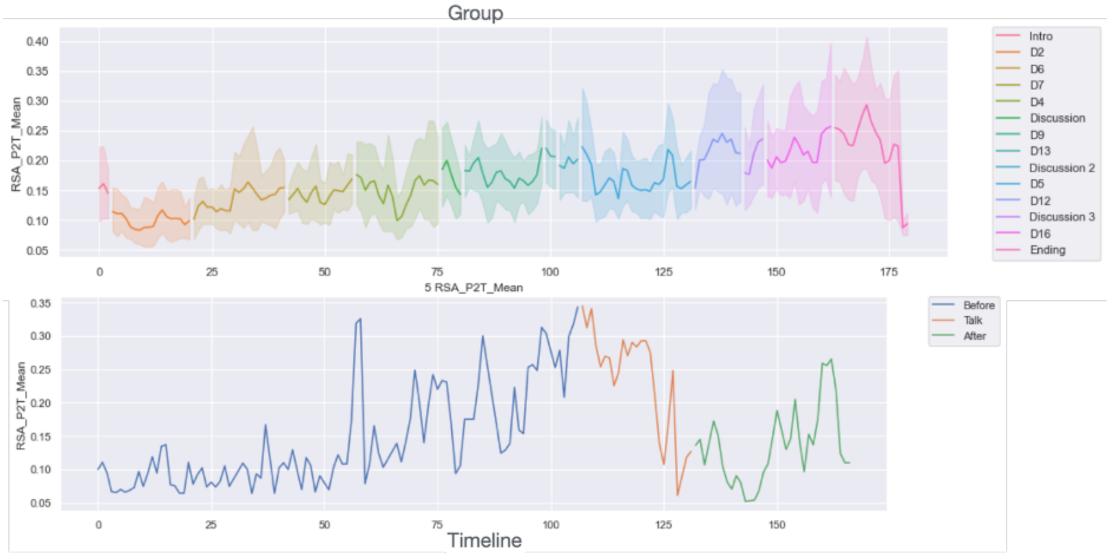

Figure 6.17 Friends Friends I: RSA feature comparison between entire group and participant wearing D5.

high the same time when group dropped and remained low. This indicated the participant with device 5 got more chilled and relaxed when he started talking and throughout his session while the rest of the group showed opposite.

Another interesting feature we found through process hrv data was the breathing rate. From figure6.18 we can see most of the y value are below 0.2 through out the Friends' Friend I event. However, most of the blue session are above 0.2 event reaching 0.25 and went back when blue session finished. The higher Y value means faster breathing rate. This mean the group are breathing way faster while participant wore 5 was sharing.

These findings were consistent with the information we gathered from traditional methods. From the questionnaire, we knew he was the participant with highest education background in that event. From the video, he often share critical opinions against other members expressed a believe against scientific methodology. There was one point, another member saying she was observing him and waiting him to criticise when someone finished sharing. These information are consist with what we can see from HRV data plotting.

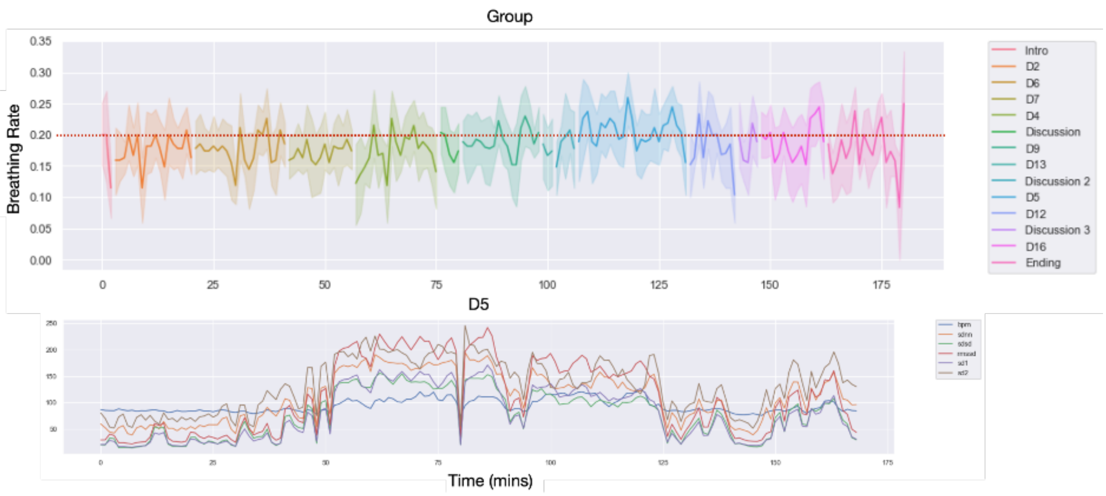

Figure 6.18 Friends' Friend I: Breathing pattern of the group (Up) and some HRV Features of D5 (bottom). In the Group breathing rate graph most of the D5 session (the blue area) are above 0.2 even reaching 0.25.

The feedback survey after the event was rather positive regarding him. 5 out of 8 people voted him as an inspiring individual and wish to know him more in future. However this does not simply suggested the conversation was successful or the voted member was popular in the group. In a follow up interview, none of the members who voted for him actually reached out for further interactions. In an interview before the event, participant wearing 5 shared his concerns about group social interaction events. He said he quite often felt being left out of the group if it's not a one to one interaction arrangement, though he seemed to be confused why people tend to not engage him in group events.

What we can tell from the feedback survey is people surely remembered him from the event. It may be more complicated when asking how people feel about him. The HRV data analysis and the video recording suggested people found him intimidating and critical in conversation. At the same time he is a very knowledgeable and informative during the event. The opposite pattern of stress levels between the group and the participant with device 5 indicate he may not be aware that he was making the group less relaxed.

6.6. Overview

Applying the proposed approach, we gave examples of how to explore the two main focuses within the study of interpersonal and social relationships. One focus is the perceptions of an individual within the social environment; as considered equally important, another is to study social reality that occurs within an individual [140]. These two perspectives helped us read and process the collected psychological data from multiple angles.

We tried different data processing methods to gather information from field recordings of social interactions. By comparing the psychophysiological data with the data gathered from traditional methods, we found the consistence of the data set. The extracted information from psychophysiological data can not be gathered by traditional measuring methods. Moreover, by combining the traditional psychometrics, we also gathered more information that may not be straight forward from traditional methods alone.

We believe adding psychophysiological data metrics to field studies can help us gather useful information to understand human behaviors and the psychological process in social interactions.

Chapter 7
Conclusions

7.1. Summary

In this thesis, we discussed the issues of current social interaction studies due to the delays between communication of different disciplines. There are delays when it comes to applying existing tools from one field to another. Even in the HCI field, most wearable devices born from the technology field lack of psychological methodology in design for use in big scale field studies. Hence, they often don't reflect well on the psychological process and are not suitable for big scale recording.

In this work we addressed a new approach of studying interpersonal relationships and social dynamics. By applying the developed system and proposed approach in daily-life scenarios outside of laboratory settings, we validated them as a possible method for social and psychological studies. We have conducted 5 workshops under different scenarios with 52 participants grouped in four relationship statuses: romantic couples, good friends, acquaintances, and new acquaintances; and four social scenarios: seminar, gatherings for meeting new friends, group board games, discussion, and a bondage practice performance. The psychophysiological data was processed in a manner of double blind.

In addition, to test the method and platform system and explore group dynamics in different social context, we conducted several large scale recordings at an international HCI conference with 80 subjects, three dance performances with 98 subjects, and two music concerts with 68 audience members. All the collected data is open source for further analysis.

7.2. Design Guidelines

For those who may be interested in conducting similar field research here are the design guidelines coming from this work. As we have discussed in chapter3 and chapter4, in order to conduct physiological tracking of social interaction in real world environment, there are particular requirements for both platform and social events.

In order to conduct successful recording for the kind of field research we focus on in this work, the tracking platform has to meet certain requirements.

- The devices have to be comfortable to wear and not distracting for participants during social interactions.
- The devices have to be cost-effective and robust for a large number of untrained participants.
- The platform has to offer access to raw data for later processing.
- The platform has to be able to function with tens or hundreds of simultaneously active recording devices.
- The platform has to be capable of managing all the incoming data and synchronize and label all the recordings as close to real time as possible.

In order to gather valid data on social interactions in real life scenarios, there are several requirements for such scenarios:

- It has to be a natural environment of social interactions, where participants are motivated to attend by themselves without interference or additional motivation from investors.
- It shall have some basic structure, so that we can arrange, group, and label the recorded data for processing.
- It has to follow a scenario where all the individuals focus on interacting with each other, with some degree of deep level communications (e.g., values, worldview, philosophy of life) or other meaningful interaction or activities.

- It shall involve a group of people and allow a one-to-one or one-to-many interaction.
- It shall last around a couple of hours or longer depending on the scenarios.

More details regarding the requirements or the platforms and the social scenarios for this work are discussed in section 4.1 and section 5.1. Besides these two fundamental designs, there are also some critical tips we learned through our experience in terms of field setting preparation and data processing.

7.2.1 Field setting

It was critical to communicate with the event organizers and get approval and support for conducting such studies with the attendees of events. It helped us to shorten the process of explanations in the field research. Thus, most of the people who joined as subjects would have some basic understanding of the field experiment procedure. Nevertheless, it is still very important to make every setting as simple as possible in order to encourage the participants to volunteer to attend the study as well as to make it easier to understand what are they expected to do during the activity. These two points are critical for getting more participants as well as valid data. We learned from the early field studies while testing the methodology 3.3.1 (UbiComp 2019 conference recording) it is important to have some kind of way to monitor and record what happened during the field research. This turned out to be one of the most important references for later data labeling and fact-checking.

7.2.2 Data

We learned two critical lessons from data collection of the field research we did in a real life environment. Since the kind of data set we recorded in this work is rather new and there is not much established ground truth for us to label and to classify the data patterns with behavior features, it is important to have information of field reality as well as a synchronized timeline for different sources of data recordings. Thus, it may be necessary to have the video recording during the field research. It can help label the data later to compare with certain behaviors or interaction patterns that happen in the field. Lacking information of field reality was one of the main problems we had when we tried to analyse the UbiComp 2019 recording. In later field research recordings, we will have our investigators set up cameras in less outstanding positions at the event locations before the participants arrive. Another important thing is to synchronize the timestamps of all the recordings as well as the video recording.

7.3. Limitations and Future Work

As we discussed before, we tried as much as we can to not interfere with the social interactions participants engage in with each other. We do not intend to manipulate any independent variable nor conclude any causal relationship in this stage. Hypothetically, the platform can be used in a field experiment which is also experimental design that the researcher control independent variables and look into cause-effect conclusions. Since we have not yet aimed for such direction, the field studies that we tested of the proposed approach did not intend to limit the extraneous variables hence low in internal validity.

Before introducing manipulation and experimental design into a research setting of this approach, we believe there are a lot of general relationships between psychophysiological data and social interaction features that need to be established.

I plan to keep working on research related to this topic and explore the potential application of this new approach even after my graduation. For instance, in the bondage performance workshop we recorded even romantic couple pairs showed that coherence and synchrony levels can be different. My hypothesis is that the synchrony level can be a predictor of relationship qualities and relationship consistency. In order to test that, we could do a follow-up study for the couples and ask them if they are still in the relationship. Also, if possible I would like to conduct more bondage workshops to gather more data and test the correlation between the synchrony levels and the nature of the pairs relationship during bondage experiences, especially to measure the relationship functioning of romantic couples with questionnaires such as The Investment Model Scale (IMS) [141]. IMS is a mean for assessing relationship quality and functioning, the IMS assesses commitment, investment, satisfaction in a relationship, and the quality of alternatives. It is also in our schedule to track couples counseling sessions and explore how that data could help to interpret close relationships. For the concert being held in April 2022, we would like to further explore the correlations between frisson sharing and the nature of the relationship of the participants.

7.4. Social Impact

The findings gathered from this research can support the design of a new system that better understands individual communication styles by analyzing social dynamics, interactions and participants' reactions in real-time. Not only psychologists can benefit from using such a system. Once it is mature enough for the general population, it can also be used to prevent poor interactions while encouraging positive ones, and facilitate higher communication satisfaction levels and improve human relationships which in return will benefit our mental health and life satisfaction in general.

7.4.1 Future Applications

The following are examples of the on-trying application cases:

Behavior Shaping

As discussed in chapter 6, what people believe is going on in a social interaction may not always be the case. With the development of this work, we believe it is possible to develop a user friendly platform and establish connections between raw psychophysiological data and real-time feedback from both the user and the group. Thus, it can help people to have a better judgement of how the conversation goes in real-time. Taking the case from section 6.5.1 as an example, the users can be more aware of how their behaviors impact the group and make changes when they start causing a lot of unwanted tension.

Counseling Therapy

Most counseling therapy relies on one-on-one interviews or sometime more than one interviewee depending on the nature of the therapy. Besides all the advantages interviews can contribute, such a method is purely qualitative. With the our approach, the therapists can get clients' feedback in more than a narrative form during the interview/talking sessions. The proposed approach may be able to add a quantitative part to current strategies in counseling therapy.

It is obvious that we can track an individual client's changes inside of one session and through out multiple sessions to offer some psychophysiological data for the therapist as a reference to understand the developments of the clients better with changes over time. Besides individual therapy, it also can offer more insights than with one client session. For instance, in sessions of couples therapy, tracking the psychophysiological data can help to see how different and similar the couple responds to the topics by looking into the synchronization levels of their data.

We have reached out to some clinical counseling communities for cooperation. Experts in the field addressed that there are not many reliable devices in the market that can acquire useful data for any meaningful interpretation. We believe the application of our hardware and software platform can contribute to changing this. With our algorithms, the physiological data collected from the clients of the psychological counselor can deliver more meaningful interpretations as a reference for the session.

Athletics

While X++ posted the information of workshops, a psychologist who joined a rock climbing community reached out to us and expressed a wish to collaborate. In the first meeting, they pointed out that for top level athletics, it's usually not the physical ability affecting their performance but the psychological process. Most of the time, the athletes will try to ask their coach or peers to give a feedback review of their performance, especially at which points they get overly nervous. A few of them record the process by video. However, it is not accurate to recall the exact mental state by memory or to paraphrase. They proposed the possibility of using psychophysiological tracking during the training to record the real-time psychological process of the athletics for later review. We plan to continue with early testing in the next year.

7.5. Conclusion

We presented a new approach for social science studies by introducing a psychophysiological sensing method into social interactions and social dynamics studies set in "real-world" environments. We found the gathered psychophysiological data from field studies reflects the data gathered from traditional methods. This result supports that we can get valid data which represents social reality. By combining this new dataset with data from traditional methods, we extract more information that was not available by traditional method alone.

As for the used platform, we concluded the requirements of the design for such psychophysiological sensing systems. In addition, we developed a smart wristband with EDA, HRV and motion sensors for psychophysiological tracking together with a software platform enabling offline and real-time data acquisition which allows 10-50 devices to simultaneously record data at the same time. We have trained machine learning models and developed algorithms to process the collected data for interpretation in the context of cognitive processes and social interactions.

In the social interaction events, we recorded 52 subjects with 143.4 participant hours. For each event, we recorded 8 to 15 people simultaneously. In musical concerts and dance performances, we recorded more than 30 participants at the same time with 149.6 participant hours in total. This scale of pyschophysiological data has not been done before. Why does data size matter? The more data we gather, the better we can pair different data patterns with behaviors as we did in the frisson study. In the future, if we can train more algorithms for behavior or emotion prediction, it can help to establish new ground truth for social interaction studies in the field as well as experimental design for causation relations of behaviors in psychological studies. All the recorded datasets are available for public and future analysis. By tracking psychophysiological data in daily-life scenarios outside of laboratory settings, we expect to validate them as a possible option for social and psychological studies.

7.6. Contribution

One of the main novelties of this work is taking mature methods or advanced tools from psychophysiology, field research, computational psychometrics, and developing an innovative new method that may cover the existing shortcomings of current approaches of social interaction research.

This thesis :

1. Provides a new approach of studying interpersonal relations, social interactions and social dynamics.
2. Builds on the field research in testing how people interact with others at individual levels and group levels.
3. Explores the use of the small wearable psychophysiological sensing devices as a data recording tool in the field studies.
4. Explores the added value of psychophysiological sensing with traditional psychometrics.
5. Explores how the psychophysiological sensing data gathering method correlates with traditional data gathering methods such as questionnaires, interviews and videos.
6. Provides data processing methods and algorithms, and data interpretation methods from the social science perspective.
7. Provides recordings and datasets available for public and future analysis.(143.4 participant-hours of recordings from 52 participants of social events, 117.6 participant-hours from dance performances, 32 participant-hours from music concerts).
8. Provides a new dimension of recordable and interpretable data from field studies in psychological studies.

References

[1] World Health Organization et al. *World health statistics 2018: monitoring health for the SDGs, sustainable development goals*. World Health Organization, 2018.

[2] Kim Bartholomew. Avoidance of intimacy: An attachment perspective. *Journal of Social and Personal relationships*, 7(2):147–178, 1990.

[3] Life in the Fast Lane (LITFL) Library. Ecg electrolyte imbalance. https://aliemcards.com/cards/ecg-electrolytes/, http://lifeinthefastlane.com/ecg-st-segment-evaluation/, September 2012. Accessed: 2021-10-25.

[4] Robert J Ellis and Julian F Thayer. Music and autonomic nervous system (dys) function. *Music perception*, 27(4):317–326, 2010.

[5] Riitta Hari and Miiamaaria V Kujala. Brain basis of human social interaction: from concepts to brain imaging. *Physiological reviews*, 89(2):453–479, 2009.

[6] C. Cannam, C. Landone, and M. Sandler. Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files. In *Proceedings of the ACM Multimedia 2010 International Conference*, pages 1467–1468, Firenze, Italy, October 2010.

[7] Pupil core - open source eye tracking platform - pupil labs. URL: https://pupil-labs.com/products/core/.

[8] E4 wristband: Real-time physiological signals: Wearable ppg, eda, temperature, motion sensors. URL: https://www.empatica.com/research/e4/.

[9] Powerlab 16/35. URL: https://www.adinstruments.com/products/powerlab/35-and-26-series.

[10] Marieke van Dooren, Joris H Janssen, et al. Emotional sweating across the body: Comparing 16 different skin conductance measurement locations. *Physiology & behavior*, 106(2):298–304, 2012.

[11] Michael E Dawson, Anne M Schell, and Diane L Filion. The electrodermal system. 2017.

[12] Scott Lilienfeld, Steven Jay Lynn, Laura Namy, Nancy Woolf, Graham Jamieson, Anthony Marks, and Virginia Slaughter. *Psychology: From inquiry to understanding 4th Edition*, volume 2. Pearson Higher Education AU, 2018.

[13] Margaret A Somerville and David Rapport. *Transdisciplinarity: Recreating integrated knowledge*. McGill-Queen's Press-MQUP, 2002.

[14] Leon Eisenberg, Terry C Pellmar, et al. Bridging disciplines in the brain, behavioral, and clinical sciences. 2000.

[15] Everett M Rogers. *Diffusion of innovations*. Simon and Schuster, 2010.

[16] Proven Models. Diffusion of innovations, 2009.

[17] Mikael Collan and Franck Tétard. Lazy user theory of solution selection. In *IADIS International Conference Cognition and Exploratory Learning in Digital Age 2007*, pages 273–278, 2007.

[18] Gheorghe Cernisov. *An unobtrusive wearable sensing platform for psychological and social dynamics*. PhD thesis, Keio UNiversity, 2020.

[19] Carl R May, Frances Mair, Tracy Finch, Anne MacFarlane, Christopher Dowrick, Shaun Treweek, Tim Rapley, Luciana Ballini, Bie Nio Ong, Anne Rogers, et al. Development of a theory of implementation and integration: Normalization process theory. *Implementation Science*, 4(1):1–9, 2009.

[20] Liz Bonnin. Bbc:the world's oddest animal couples. https://www.bbc.co.uk/programmes/b03jp46s, November 2013.

[21] Laura Helena Andrade, J Alonso, Z Mneimneh, JE Wells, A Al-Hamzawi, G Borges, E Bromet, Ronny Bruffaerts, G De Girolamo, R De Graaf, et al. Barriers to mental health treatment: results from the who world mental health surveys. *Psychological medicine*, 44(6):1303–1317, 2014.

[22] Morris A Okun, William A Stock, Marilyn J Haring, and Robert A Witter. Health and subjective well-being: A meta-analyis. *The International journal of aging and human development*, 19(2):111–132, 1984.

[23] Sonja Lyubomirsky, Laura King, and Ed Diener. The benefits of frequent positive affect: Does happiness lead to success? *Psychological bulletin*, 131(6):803, 2005.

[24] LeeAnne Harker and Dacher Keltner. Expressions of positive emotion in women's college yearbook pictures and their relationship to personality and life outcomes across adulthood. *Journal of personality and social psychology*, 80(1):112, 2001.

[25] Gary N Marks and Nicole Fleming. Influences and consequences of well-being among australian young people: 1980–1995. *Social indicators research*, 46(3):301–323, 1999.

[26] Karen L Siedlecki, Timothy A Salthouse, Shigehiro Oishi, and Sheena Jeswani. The relationship between social support and subjective well-being across age. *Social indicators research*, 117(2):561–576, 2014.

[27] Auguste Comte. *System of Positive Polity: Theory of the future of man, with an appendix consisting of Early essays on social philosophy*, volume 4. Longmans, Green, and Company, 1877.

[28] Gary R VandenBos. *APA dictionary of psychology.* American Psychological Association, 2007.

[29] Kurt Lewin. Frontiers in group dynamics: Ii. channels of group life; social planning and action research. *Human relations*, 1(2):143–153, 1947.

[30] Steven N Durlauf and H Peyton Young. *Social dynamics*, volume 4. Mit Press, 2004.

[31] Pitirim Aleksandrovich Sorokin. Contemporary sociological theories. 1928.

[32] William Little, Ron McGivern, and Nathan Kerins. *Introduction to sociology-2nd Canadian edition*. BC Campus, 2016.

[33] Susan Curtiss. *Genie: a psycholinguistic study of a modern-day wild child*. Academic Press, 1977.

[34] Russ Rymer. Genie: A scientific tragedy. 1993.

[35] Stephen D. Krashen. Lateralization, language learning, and the critical period: Some new evidence. *Language Learning*, 23(1):63–74, 1973.

[36] Sara Shahrestani, Elizabeth M Stewart, Daniel S Quintana, Ian B Hickie, and Adam J Guastella. Heart rate variability during adolescent and adult social interactions: A meta-analysis. *Biological psychology*, 105:43–50, 2015.

[37] Masako Ishii-Kuntz. Social interaction and psychological well-being: Comparison across stages of adulthood. *The International Journal of Aging and Human Development*, 30(1):15–36, 1990.

[38] World Health Organization et al. *World health statistics 2021: monitoring health for the SDGs, sustainable development goals*. World Health Organization, 2021.

[39] Sonja Lyubomirsky, Kennon M Sheldon, and David Schkade. Pursuing happiness: The architecture of sustainable change. *Review of general psychology*, 9(2):111–131, 2005.

[40] Ed Diener and Martin EP Seligman. Very happy people. *Psychological science*, 13(1):81–84, 2002.

[41] Joshua Wolf Shenk. What makes us happy. *The Atlantic*, 48, 2009.

[42] Vaillant Ge. Mukamal k. successful aging. *American Journal of Psychiatry*, 158(6):839–847, 2001.

[43] James A Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161, 1980.

[44] John Cassel. The contribution of the social environment to host resistance: the fourth wade hampton frost lecture. *American journal of epidemiology*, 104(2):107–123, 1976.

[45] Sidney Cobb. Social support as a moderator of life stress. *Psychosomatic medicine*, 1976.

[46] Lisa M Diamond. Contributions of psychophysiology to research on adult attachment: Review and recommendations. *Personality and Social Psychology Review*, 5(4):276–295, 2001.

[47] Bert N Uchino, John T Cacioppo, and Janice K Kiecolt-Glaser. The relationship between social support and physiological processes: a review with emphasis on underlying mechanisms and implications for health. *Psychological bulletin*, 119(3):488, 1996.

[48] Mary Main, Nancy Kaplan, and Jude Cassidy. Security in infancy, childhood, and adulthood: A move to the level of representation. *Monographs of the society for research in child development*, pages 66–104, 1985.

[49] R Chris Fraley, Amanda M Vicary, Claudia Chloe Brumbaugh, and Glenn I Roisman. Patterns of stability in adult attachment: an empirical test of two models of continuity and change. *Journal of personality and social psychology*, 101(5):974, 2011.

[50] Ross A Thompson, Jeffry A Simpson, and Lisa J Berlin. Attachment theory in the 21st century. *Attachment: The Fundamental Questions*, page 7, 2021.

[51] Cindy Hazan and Phillip Shaver. Romantic love conceptualized as an attachment process. *Journal of personality and social psychology*, 52(3):511, 1987.

[52] Kim Bartholomew and Leonard M Horowitz. Attachment styles among young adults: a test of a four-category model. *Journal of personality and social psychology*, 61(2):226, 1991.

[53] Mary A Snydersmith and John T Cacioppo. Parsing complex social factors to determine component effects: I. autonomic activity and reactivity as a

function of human association. *Journal of Social and Clinical Psychology*, 11(3):263–278, 1992.

[54] David Sheffield and Douglas Carroll. Social support and cardiovascular reactions to active laboratory stressors. *Psychology & Health*, 9(4):305–316, 1994.

[55] Anne M Fontana, Tracey Diegnan, Angela Villeneuve, and Stephen J Lepore. Nonevaluative social support reduces cardiovascular reactivity in young women during acutely stressful performance situations. *Journal of Behavioral Medicine*, 22(1):75–91, 1999.

[56] Clemens Kirschbaum, Thomas Klauer, Sigrun-Heide Filipp, and Dirk Helmut Hellhammer. Sex-specific effects of social support on cortisol and subjective responses to acute psychological stress. *Psychosomatic medicine*, 57(1):23–31, 1995.

[57] Rajiv S Jhangiani, I-Chant A Chiang, Carrie Cuttler, Dana C Leighton, et al. *Research methods in psychology.* Kwantlen Polytechnic University, 2019.

[58] Michael Angrosino. *Doing ethnographic and observational research.* Sage, 2007.

[59] Gerald C Davison and Arnold A Lazarus. Clinical case studies are important in the science and practice of psychotherapy. 2007.

[60] Paul C Cozby, Scott Bates, Chris Krageloh, Philippe Lacherez, and Dirk Van Rooy. *Methods in behavioral research.* Mayfield publishing company Houston, TX, 1977.

[61] John F Edens, Jacqueline K Buffington, Tara L Tomicic, and Brandon D Riley. Effects of positive impression management on the psychopathic personality inventory. *Law and Human Behavior*, 25(3):235–256, 2001.

[62] Delroy L Paulhus. Measurement and control of response bias. 1991.

[63] Richard E Nisbett and Timothy D Wilson. Telling more than we can know: Verbal reports on mental processes. *Psychological review*, 84(3):231, 1977.

[64] Thomas F Oltmanns and Eric Turkheimer. Person perception and personality pathology. *Current Directions in Psychological Science*, 18(1):32–36, 2009.

[65] Oliver P John and Richard W Robins. Accuracy and bias in self-perception: individual differences in self-enhancement and the role of narcissism. *Journal of personality and social psychology*, 66(1):206, 1994.

[66] W Keith Campbell and Joshua D Miller. *The handbook of narcissism and narcissistic personality disorder*. Wiley Online Library, 2011.

[67] John T Cacioppo, Louis G Tassinary, and Gary Berntson. *Handbook of psychophysiology*. Cambridge university press, 2007.

[68] Paul William Harkins, Walther Riese, et al. *Galen on the Passions and Errors of the Soul*. The Ohio State University Press, 1963.

[69] Mathias Benedek and Christian Kaernbach. A continuous measure of phasic electrodermal activity. *Journal of neuroscience methods*, 190(1):80–91, 2010.

[70] Christian Tronstad, Gaute E Gjein, Sverre Grimnes, Ørjan G Martinsen, Anne-Lene Krogstad, and Erik Fosse. Electrical measurement of sweat activity. *Physiological measurement*, 29(6):S407, 2008.

[71] Carl Gustav Jung. *Studies in word-association*. Taylor & Francis, 1969.

[72] TW Picton, I Martin, and PH Venables. Techniques in psychophysiology. 1980.

[73] John L Andreassi. *Psychophysiology: Human behavior and physiological response*. Psychology Press, 2010.

[74] Wolfram Boucsein. Electrodermal activity: Springer science & business media. *Broek, EL vd, Schut, MH, Westerink, JHDM, Herk, J. v., & Tuinenbreijer, K*, 2012.

[75] Peter J Lang, Mark K Greenwald, Margaret M Bradley, and Alfons O Hamm. Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, 30(3):261–273, 1993.

[76] Fred Shaffer and Jay P Ginsberg. An overview of heart rate variability metrics and norms. *Frontiers in public health*, page 258, 2017.

[77] George E Billman. Heart rate variability–a historical perspective. *Frontiers in physiology*, 2:86, 2011.

[78] Kwang-Ho Choi, Junbeom Kim, O Sang Kwon, Min Ji Kim, Yeon Hee Ryu, and Ji-Eun Park. Is heart rate variability (hrv) an adequate tool for evaluating human emotions?–a focus on the use of the international affective picture system (iaps). *Psychiatry research*, 251:192–196, 2017.

[79] Julian F Thayer and Richard D Lane. A model of neurovisceral integration in emotion regulation and dysregulation. *Journal of affective disorders*, 61(3):201–216, 2000.

[80] Stephen W Porges. The polyvagal perspective. *Biological psychology*, 74(2):116–143, 2007.

[81] Fumihiko Yasuma and Jun-ichiro Hayano. Respiratory sinus arrhythmia: why does the heartbeat synchronize with respiratory rhythm? *Chest*, 125(2):683–690, 2004.

[82] Emily A Butler, Frank H Wilhelm, and James J Gross. Respiratory sinus arrhythmia, emotion, and emotion regulation during social interaction. *Psychophysiology*, 43(6):612–622, 2006.

[83] Michael Thompson, Lynda Thompson, and Andrea Reid-Chung. Combining loreta z-score neurofeedback with heart rate variability training. *Z Score Neurofeedback: Clinical Applications*, 2014.

[84] Przemysław Guzik, Jarosław Piskorski, Tomasz Krauze, Raphael Schneider, Karel H Wesseling, Andrzej Wykretowicz, and Henryk Wysocki. Correlations between poincaré plot and conventional heart rate variability parame-

ters assessed during paced breathing. *The Journal of Physiological Sciences*, pages 0702020009–0702020009, 2007.

[85] Ali Motie-Nasrabadi, Soroor Behbahani, and Nader Jafarnia Dabanloo. Ictal heart rate variability assessment with focus on secondary generalized and complex partial epileptic seizures. *Advances in Bioresearch*, 4(1), 2013.

[86] Theodore Beauchaine. Vagal tone, development, and gray's motivational theory: Toward an integrated model of autonomic nervous system functioning in psychopathology. *Development and psychopathology*, 13(2):183–214, 2001.

[87] Stephen W Porges. Orienting in a defensive world: Mammalian modifications of our evolutionary heritage. a polyvagal theory. *Psychophysiology*, 32(4):301–318, 1995.

[88] Clemens Kirschbaum, Karl-Martin Pirke, and Dirk H Hellhammer. The 'trier social stress test'–a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, 28(1-2):76–81, 1993.

[89] Beate Ditzen, Inga D Neumann, Guy Bodenmann, Bernadette von Dawans, Rebecca A Turner, Ulrike Ehlert, and Markus Heinrichs. Effects of different kinds of couple interaction on cortisol and heart rate responses to stress in women. *Psychoneuroendocrinology*, 32(5):565–574, 2007.

[90] Yvonne Rogers and Paul Marshall. Research in the wild. *Synthesis Lectures on Human-Centered Informatics*, 10(3):i–97, 2017.

[91] Nickola C Overall, Jeffry A Simpson, and Helena Struthers. Buffering attachment-related avoidance: Softening emotional and behavioral defenses during conflict discussions. *Journal of personality and social psychology*, 104(5):854, 2013.

[92] Pietro Cipresso, Aleksandar Matic, Dimitris Giakoumis, and Yuri Ostrovsky. Advances in computational psychometrics, 2015.

[93] Alina A von Davier. Computational psychometrics in support of collaborative educational assessments, 2017.

[94] Robert J Mislevy. Next generation learning and assessment: What, why and how. In *Computational Psychometrics: New Methodologies for a New Generation of Digital Learning and Assessment*, pages 9–24. Springer, 2021.

[95] Shoya Ishimaru, Kai Kunze, Yuji Uema, Koichi Kise, Masahiko Inami, and Katsuma Tanaka. Smarter eyewear: using commercial eog glasses for activity recognition. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pages 239–242, 2014.

[96] Dingding Zheng, Laura Lugaresi, George Chernyshov, Benjamin Tag, Masa Inakage, and Kai Kunze. Wearable aura: an interactive projection on personal space to enhance communication. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, pages 141–144. ACM, 2017.

[97] Laura Lugaresi, Kaiyuan Lin, and Dingding Zheng. Wearable aura: Interactive personal projection to bring people closer. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–6, 2018.

[98] Fangqing Zhengren, George Chernyshov, Dingding Zheng, and Kai Kunze. Cognitive load assessment from facial temperature using smart eyewear. In *Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, pages 657–660. ACM, 2019.

[99] Thomas K Ferris and Nadine B Sarter. Cross-modal links among vision, audition, and touch in complex environments. *Human Factors*, 50(1):17–26, 2008.

[100] Vic Tandy and Tony R Lawrence. The ghost in the machine. *Journal-Society for Psychical Research*, 62:360–364, 1998.

[101] Kirill Ragozin, Dingding Zheng, George Chernyshov, and Danny Hynds. Sophroneo: Fear not. a vr horror game with thermal feedback and physio-

logical signal loop. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–6, 2020.

[102] Kirill Ragozin, George Chernyshov, Dingding Zheng, Danny Hynds, Jianing Zhao, Kouta Minamizawa, and Kai Kunze. Sophroneo: Fear not. a vr horror game with thermal feedback and physiological signal loop. In *ACM SIGGRAPH 2020 Immersive Pavilion*, pages 1–2. 2020.

[103] Kouta Minamizawa, Yasuaki Kakehi, Masashi Nakatani, Soichiro Mihara, and Susumu Tachi. Techtile toolkit: a prototyping tool for designing haptic media. 08 2012. `doi:10.1145/2343456.2343478`.

[104] Yulan Ju, Dingding Zheng, Danny Hynds, George Chernyshov, Kai Kunze, and Kouta Minamizawa. Haptic empathy: Conveying emotional meaning through vibrotactile feedback. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–7, 2021.

[105] Jiawen Han, George Chernyshov, Dingding Zheng, Peizhong Gao, Takuji Narumi, Katrin Wolf, and Kai Kunze. Sentiment pen: recognizing emotional context based on handwriting features. In *Proceedings of the 10th Augmented Human International Conference 2019*, page 24. ACM, 2019.

[106] Scott Bannister and Tuomas Eerola. Suppressing the chills: Effects of musical manipulation on the chills response. *Frontiers in psychology*, 9:2046, 2018.

[107] Ming-Zher Poh. *Continuous assessment of epileptic seizures with wrist-worn biosensors*. PhD thesis, Massachusetts Institute of Technology, 2011.

[108] Lin Shu, Jinyan Xie, Mingyue Yang, Ziyi Li, Zhenqi Li, Dan Liao, Xiangmin Xu, and Xinyi Yang. A review of emotion recognition using physiological signals. *Sensors*, 18(7):2074, 2018.

[109] Dominique Makowski, Tam Pham, Zen J. Lau, Jan C. Brammer, François Lespinasse, Hung Pham, Christopher Schölzel, and S. H. Annabel Chen. Neurokit2: A python toolbox for neurophysiological signal processing. *Behavior Research Methods*, Feb 2021. URL: `https://doi.org/10.3758/s13428-020-01516-y`, `doi:10.3758/s13428-020-01516-y`.

[110] Moe Sugawa, Taichi Furukawa, George Chernyshov, Danny Hynds, Jiawen Han, Marcelo Padovani, Dingding Zheng, Karola Marky, Kai Kunze, and Kouta Minamizawa. Boiling mind: Amplifying the audience-performer connection through sonification and visualization of heart and electrodermal activities. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*, pages 1–10, 2021.

[111] Zhuoqi Fu, Jiawen Han, Dingding Zheng, Moe Sugawa, Taichi Furukawa, Chernyshov George, Hynds Danny, Padovani Marcelo, Marky Karola, Kouta Minamizawa, et al. Boiling mind-a dataset of physiological signals during an exploratory dance performance. In *Augmented Humans Conference 2021*, pages 301–303, 2021.

[112] Y He, D Zheng, G Chernyshov, R Thomsen, J Han, D Hynds, Y S Pai, K Kunze, and K Minamizawa. Frisson waves: Sharing frisson to create collective empathetic experiences for music performances. In *2021 IEEE World Haptics Conference (WHC)*, pages 591–591. IEEE, 2021.

[113] Elise S Dan-Glauser and Klaus R Scherer. The geneva affective picture database (gaped): a new 730-picture database focusing on valence and normative significance. *Behavior research methods*, 43(2):468, 2011.

[114] Eckhard H Hess and James M Polt. Pupil size as related to interest value of visual stimuli. *Science*, 132(3423):349–350, 1960.

[115] Bruno Laeng, Sylvain Sirois, and Gustaf Gredebäck. Pupillometry: A window to the preconscious? *Perspectives on psychological science*, 7(1):18–27, 2012.

[116] Alexandros Kafkas and Daniela Montaldi. Recognition memory strength is predicted by pupillary responses at encoding while fixation patterns distinguish recollection from familiarity. *Quarterly journal of experimental psychology*, 64(10):1971–1989, 2011.

[117] Eric Ed Granholm and Stuart R Steinhauer. Pupillometric measures of cognitive and emotional processes. *International Journal of Psychophysiology*, 2004.

[118] Hiram E Fitzgerald. Autonomic pupillary reflex activity during early infancy and its relation to social and nonsocial visual stimuli. *Journal of Experimental Child Psychology*, 6(3):470–482, 1968.

[119] Gary Aston-Jones and Jonathan D Cohen. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.*, 28:403–450, 2005.

[120] Peter Dayan and Angela J Yu. Phasic norepinephrine: a neural interrupt signal for unexpected events. *Network: Computation in Neural Systems*, 17(4):335–350, 2006.

[121] Margaret M Bradley, Laura Miccoli, Miguel A Escrig, and Peter J Lang. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45(4):602–607, 2008.

[122] Robert R Henderson, Margaret M Bradley, and Peter J Lang. Emotional imagery and pupil diameter. *Psychophysiology*, 55(6):e13050, 2018.

[123] Cindy HAZAN. Conceptualizing romantic love as an attachment process. *Journal of personality and social psychology*, 52:511–524, 1987.

[124] Nancy L Collins. Working models of attachment: Implications for explanation, emotion, and behavior. *Journal of personality and social psychology*, 71(4):810, 1996.

[125] Nancy L Collins and Stephen J Read. Adult attachment, working models, and relationship quality in dating couples. *Journal of personality and social psychology*, 58(4):644, 1990.

[126] Mary D Salter Ainsworth. The development of infant-mother attachment. In *In Thebeginning*, pages 133–143. Columbia University Press, 1982.

[127] Mary DS Ainsworth, Mary Blehar, Everett Waters, and Sally Wall. Patterns of attachment, 1978.

[128] Eleanor E Maccoby. *Social development: Psychological growth and the parent-child relationship*. New York: Harcourt Brace Jovanovich, 1980.

[129] Marc B Levy and Keith E Davis. Lovestyles and attachment styles compared: Their relations to each other and to various relationship characteristics. *Journal of social and Personal Relationships*, 5(4):439–471, 1988.

[130] R Chris Fraley, Marie E Heffernan, Amanda M Vicary, and Claudia Chloe Brumbaugh. The experiences in close relationships—relationship structures questionnaire: A method for assessing attachment orientations across relationships. *Psychological assessment*, 23(3):615, 2011.

[131] Ulric Neisser and Nicole Harsch. Phantom flashbulbs: False recollections of hearing the news about challenger. 1992.

[132] Heike Schmolck, Elizabeth A Buffalo, and Larry R Squire. Memory distortions develop over time: Recollections of the oj simpson trial verdict after 15 and 32 months. *Psychological Science*, 11(1):39–45, 2000.

[133] Ulrike Rimmele, Lila Davachi, and Elizabeth A Phelps. Memory for time and place contributes to enhanced confidence in memories for emotional events. *Emotion*, 12(4):834, 2012.

[134] Roger Brown and James Kulik. Flashbulb memories. *Cognition*, 5(1):73–99, 1977.

[135] Jennifer M Talarico and David C Rubin. Flashbulb memories are special after all; in phenomenology, not accuracy. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 21(5):557–578, 2007.

[136] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms

for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi:10.1038/s41592-019-0686-2.

[137] Fred Shaffer, Rollin McCraty, and Christopher L Zerr. A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. *Frontiers in psychology*, 5:1040, 2014.

[138] Luca Citi, Emery N Brown, and Riccardo Barbieri. A real-time automated point-process method for the detection and correction of erroneous and ectopic heartbeats. *IEEE transactions on biomedical engineering*, 59(10):2828–2837, 2012.

[139] Toni Giorgino. Computing and visualizing dynamic time warping alignments in r: the dtw package. *Journal of statistical Software*, 31:1–24, 2009.

[140] Wendi L Gardner, Shira Gabriel, and Amanda B Diekman. Interpersonal processes. *Handbook of psychophysiology*, pages 643–664, 2000.

[141] Caryl E Rusbult, John M Martz, and Christopher R Agnew. The investment model scale: Measuring commitment level, satisfaction level, quality of alternatives, and investment size. *Personal relationships*, 5(4):357–387, 1998.

Appendices

A. THE RELATIONSHIPS QUESTIONNAIRE (RQ)

1. Following are four general relationship styles that people often report. Please choose the letter corresponding to the style that best describes you or is closest to the way you are.

A. It is easy for me to become emotionally close to others. I am comfortable depending on them and having them depend on me. I don't worry about being alone or having others not accept me.

B. I am uncomfortable getting close to others. I want emotionally close relationships, but I find it difficult to trust others completely, or to depend on them. I worry that I will be hurt if I allow myself to become too close to others.

C. I want to be completely emotionally intimate with others, but I often find that others are reluctant to get as close as I would like. I am uncomfortable being without close relationships, but I sometimes worry that others don't value me as much as I value them.

D. I am comfortable without close emotional relationships. It is very important to me to feel independent and self-sufficient, and I prefer not to depend on others or have others depend on me.

Now please rate each of the relationship styles above to indicate how well or poorly each description corresponds to your general relationship style. Please think about all your relationships (past and present) and respond in terms of how you generally feel in these relationships. If you have never been involved in a romantic relationship, answer in terms of how you think you would feel.

Please use the scale below by choosing a number between 1 and 7 for each statement.

2. It is easy for me to become emotionally close to others. I am comfortable depending on them and having them depend on me. I don't worry about being alone or having others not accept me.

1 2 3 4 5 6 7

3. I am uncomfortable getting close to others. I want emotionally close relationships, but I find it difficult to trust others completely, or to depend on them. I worry that I will be hurt if I allow myself to become too close to others.

1 2 3 4 5 6 7

4. I want to be completely emotionally intimate with others, but I often find that others are reluctant to get as close as I would like. I am uncomfortable being without close relationships, but I sometimes worry that others don't value me as much as I value them.

1 2 3 4 5 6 7

5. I am comfortable without close emotional relationships. It is very important to me to feel independent and self-sufficient, and I prefer not to depend on others or have others depend on me.

1 2 3 4 5 6 7