Doctoral Dissertation
Academic Year 2022

From Physiology to Group Dynamics: A Practical Framework for Physiological Data Analysis

Keio University
Graduate School of Media Design

Jiawen Han

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

Jiawen Han

Dissertation Advisory Committee:

Professor Kai Kunze	(Principal Advisor)
Professor Keiko Okawa	(Co-Advisor)
Professor Akira Kato	(Co-Advisor)

Doctoral Dissertation Review Committee:

Professor Keiko Okawa	(Chair)
Professor Akira Kato	(Member)
Professor Jamie A.Ward	(Member, Goldsmiths University of London)
Professor Karola Marky	(Member, Ruhr University Bochum)

Abstract of Doctoral Dissertation of Academic Year 2022

From Physiology to Group Dynamics: A Practical Framework for Physiological Data Analysis

Category: Science / Engineering

Summary

The advent of wearable sensing technologies and neuroscience basis proving the uses of physiological data as social signals bring possibilities for understanding group dynamics through physiological data analysis. However, for researchers without physiological data processing knowledge, it is usually complex to use physiological data as measurements to investigate in-the-wild group dynamics. It is more challenging for practitioners in the HCI field to apply physiological data analysis to sensor-based interactions during large-scale group events. To bridge this gap, this thesis describes a practical framework for physiological data analysis based on standard physiological data processing procedures and centers on the concept of entrainment to understand and augment group dynamics. Entrainment at the physiological level can be used as an objective measure of internal processes accompanying empathic interactions related to group cohesion, connectedness, and engagement. On the other hand, biofeedback that reveals the hidden dynamics can also influence live experience of individuals and the collective reaction of a group. Therefore, we adopt this concept to help researchers and practitioners in the HCI field transform physiological data into research insights to augment group interactions.

There are three main contributions of this thesis. Firstly, this thesis elaborates on the proposed analysis framework by describing the key steps when conducting offline analysis and real-time analysis. In this thesis, offline analysis is defined as the analysis conducted on the recorded datasets to understand collective experience in group events. While real-time analysis is defined as analysis algorithms implemented in biofeedback systems that require a relatively short response time to trigger feedback. This thesis explores how to adopt the concept of entrainment in the five essential steps of physiological data analysis: data collection, preprocessing, feature extraction, analysis, and interpretation.

Secondly, we applied this analysis framework and explored blood volume pulse (BVP) data and electrodermal activity (EDA) data collected from participants in group events such as social games, online lectures, and performances. We found several explainable features that could be used to quantify group dynamics. For example, pNN50 is an HRV feature that is closely related to the activation of the parasympathetic nervous system (PSNS). Therefore, we could adopt pNN50 as an indicator for relaxation. Moreover, this thesis summarizes approaches to analyzing and interpreting physiological data in a collective manner. One method is computing the trends of physiological features and mapping them to the development of group events. With annotations, such as the unfolding of the storyline and notable moments, this method could provide a holistic view of the in-the-wild experience. Another method is calculating the similarity between each pair in the group as pair-wise entrainment and comparing between groups by applying statistical analysis. This method may need additional information to distinguish different groups before comparison. We also publicized the physiological dataset collected with offline analysis sample codes that are freely available for the HCI community.

Thirdly, this thesis presents the methodology of developing real-time analysis to detect and share physiological experience in the group. The process of collecting labeled data in a lab study, training a machine learning detection model and implementing the detection model in a real-life biofeedback system is described. This opens up a potential direction to augment group interaction by recognizing and influencing physiological experience.

Keywords:

physiology, group dynamics, wearable sensing, biofeedback, in-the-wild experience, practice-led research

Keio University Graduate School of Media Design

Jiawen Han

Contents

List of Figures

List of Tables

Acknowledgements

First, I want to thank my principal supervisor Kai Kunze for his valuable advice, support, and encouragement. He is always supportive and excited when discussing research ideas and provides me great opportunities to talk with outstanding scholars. I want to thank my other supervisors who have helped me organize my research ideas with constructive feedback: Keiko Okawa, Akira Kato, Jamie A. Ward, and Karola Marky. I think their expert knowledge, passionate working styles, and respectable personalities will always inspire me to challenge higher goals. I also want to thank Shaoke Zhang who was one of my undergraduate professors. He firstly introduced amazing research works published at the CHI conference, which led me to the interesting HCI world.

Many thanks to my talented colleagues in Geist who built up amazing prototypes and shared enlightening ideas: Gheorghe Chernisov, Dingding Zheng, Kanyu Chen, Zhuoqi Fu, other Geist members, and our secretary Mio Sugimoto. I also want to express my great thanks to other collaborators who worked on challenging tasks together with me: Ziyue Wang, Chi-Lan Yang, Reiya Horii, and Danny Hynds. Especially, I would like to thank Moe Sugawa and Yan He, who invited me to join their amazing projects that enlightened my research direction.

I want to thank my family, my friends, and my haru puppy. It is because of their love that I could survive the anxiety and self-doubt. My parents are always supportive when I am chasing my dream and making decisions. My husband has always accompanied me with his limitless understanding and practical advice when I was learning, struggling, and challenging. My friends are always cheering me up when I am stuck and down. Even a short coffee time or a remote call really healed me. Maybe, I would also say thanks to myself who decided to start this unforgettable journey and did not give up. This research journey helped me gained my power, autonomy, and self-confidence.

Finally, I am grateful that Keio University offered me the Design the Future

Award. This scholarship has not only alleviated my financial burden but also provided me with great opportunities to talk with excellent researchers. This honor will always encourage me in the future.

Chapter 1
Introduction and Background

1.1. Motivation

Emotion has an essential role in human behavior also relating to cognition and perception [1]. However, it is not always easy for people to catch and exchange affective feedback considering the ambiguity of emotion. One of the most established perspectives to quantify emotion starts from the physiology of emotion and investigates emotion-specific autonomic nervous system (ANS) activation [2, 3]. Recently the development of physiological sensing has enabled understanding from psychophysiological and neurophysiological perspectives related to ANS. ANS has been proven to influence human experience not only at the intrapersonal level but also "across-subject" in terms of its externally responsive feature. This suggests that ANS responding could be used to understand group dynamics beyond individual emotional experiences [4].

Previous work also explored sensing live group physiology using a variety of different sensor technologies, including electrodermal activity (EDA) [5–7], heart rate variability (HRV) [8–10], electroencephalogram (EEG) [11], and body movement [12–14] (see Table 2.1in page 12). However, there is no scientific consensus on how to link individual physiological responses to an affective experience in group dynamics. My research work is to explore the methodologies of analyzing physiological data beyond individual experiences centering around the concept of entrainment, which will be explained below.

During group interactions, social signals tend to be communicated through unconscious behaviors [15, 16] and group dynamics are usually reflected by the unintentional coordination [16]. The process of coordination is referred to as "entrainment". Borrowed from physics [17], entrainment describes the process by which independent rhythmical systems interact with each other so that they adjust themselves and eventually become rhythmically coupled [8, 18]. Further, this notion was enlarged to dynamic coordinated behaviors and even internal physiological activities related to the "bond" between human beings [19].

Unlike behavioral signs, physiological entrainment is unable to be naturally observed by the naked eye. However, the advent of wearable sensing devices has enabled capturing physiological signals at a relatively unobtrusive level [20]. The application of time series analysis also provided more insights into evaluating physiological entrainment and proved its positive effect on understanding group cohesion and team trust [21]. As an unconscious and "invisible" coordination, physiological entrainment can reflect the underlying bodily and neuronal dynamics during group interactions. On the other hand, biofeedback that reveals the hidden dynamics can also influence each individual's live experience and the collective reaction of a group [22–26]. Therefore, measurements centering the concept of physiological entrainment could bring more possibilities to augment group dynamics.

To investigate the potential of this concept, we extended physiological data analysis from lab studies to in-the-wild group experience. According to Benford et al., "in-the-wild" research in the HCI field could be defined as the fusion of computing technology and public artistic projects (including performance) in the sense of engaging "real" users with emerging technologies in real settings under demanding conditions of actual use, as opposed to the more constrained "lab" environment [27]. Understanding and augmenting group dynamics during in-the-wild experience could benefit both practitioners and researchers. For practitioners and artists, sensing and analyzing live group experience during in-the-wild events could create novel interactions and gain insights into how audience experience their works under a reflexive process. For researchers, the opportunity to gather and analyze in-the-wild dataset could reveal natural responses hidden in the lab environments [27, 28]. Accordingly, it is quite challenging to understand this mobile, ubiquitous, and complex in-the-wild experience [27]. This thesis aims to explore methodologies to quantify, interpret, and augment in-the-wild group dynamics.

1.2. Research Background

A conventional way to investigate the subjective experience in group dynamics is measuring engagement and participation by using a combination of observations during the group interactions and the collection of responses via surveys afterwards. However, this type of approaches is highly subjective and surveys might be difficult to deploy in some scenarios, such as artistic performances because audience members tend to think it is market-related [29]. Lying in the Human-Centred Computing [30] and Human-Computer Integration paradigms [31–35], measuring people's unconsciously generated physiological data with wearable sensing technologies have become complementary methods (see Table 2.1). Yet, it is quite challenging to specify the relationship between sensor inputs and practical insights. One reason for this is that it is not visually apparent to match raw and continuous data to high-level concepts, such as engagement in group interactions. Another reason is that existing tools for analyzing sensor data are usually designed for professional engineers and scientists [36]. Therefore, we propose a practical analysis framework centering around the notion of entrainment to help researchers in the HCI field and practitioners working on sensor-based interactions. Physiological entrainment is an essential part of shared experiences and could be used as an objective measure of internal processes accompanying empathic interactions [4,37]. This framework focuses on the coordination and similarity among the patterns in the physiological data of multiple people, namely the "bond" within the group [19]. Considering the availability of real-time monitoring physiological data, the proposed framework consists of two components (see Figure 1.1):

- An offline analysis component where we propose a universal analysis flow with interpretation perspectives to investigate subjective experience in certain groups. And how the methodology could be applied to different scenarios.

- A real-time analysis component where we propose algorithms to detect and characterize the collective physiological flow and critical physiological events for augmenting subjective experience in certain groups. The algorithms could be applied to multiple people in real-time and support biofeedback practices.

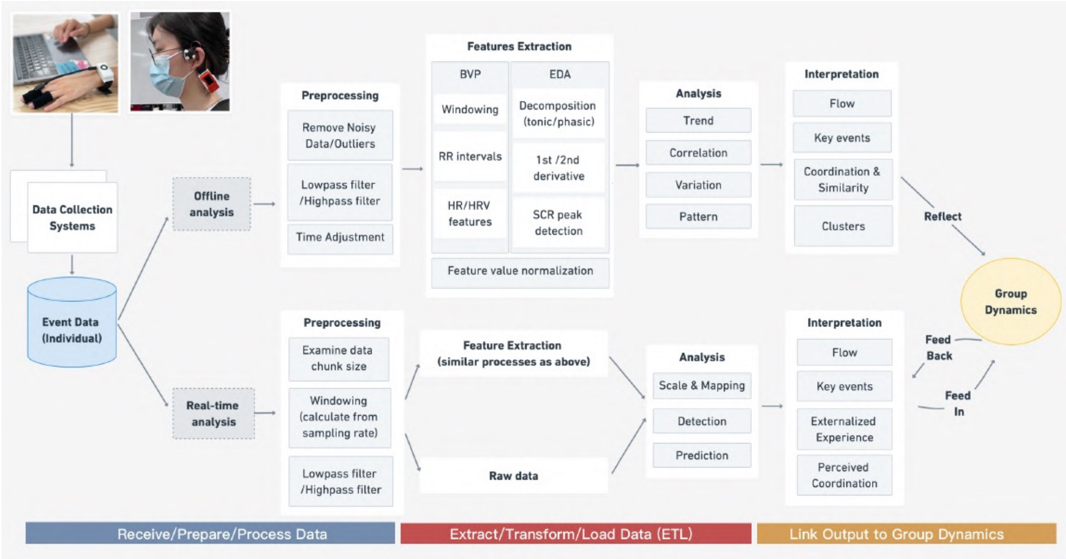

Figure 1.1 Proposed physiological data analysis framework to understand and augment group dynamics.

For each component, we either contribute new algorithms or suggest new ways to make use of existing analytical methodologies to understand group dynamics. To explain and complete the proposed framework, we describe the process and research insights by analyzing physiological data (focusing on BVP and EDA data in my research) collected from groups in different scenarios.

1.3. Research Questions

The specific research questions we would like to answer to support and complete the proposed framework are as follows:

1. How can we use the concept of entrainment to improve understanding of group dynamics by physiological data?
   (a) How can the proposed offline analysis be used to quantify group dynamics beyond individual subjective experience?

   (b) Which aspects can research insights acquired in offline analysis imply real-time analysis in biofeedback systems?

2. How can we use the concept of entrainment to improve augmenting group dynamics by physiological data?
   (a) How can the proposed framework for physiological data analysis be applied to real-life biofeedback systems?
   (b) What effects do the biofeedback systems embedded with the proposed real-time analysis bring to group interactions?

3. How to integrate the proposed framework with practical goals during interdisciplinary collaborations?

1.4. Research Contributions

The final contributions of my research are as follows:

- A physiological data analysis framework to understand and augment group dynamics centering around the concept of entrainment that could be applied in both offline analysis afterward and real-time analysis during biofeedback practices. The framework extends physiological data analysis from dyadic level to group level and from lab studies to in-the-wild group activities.

- Reproducible analysis processes and research insights from the offline analysis on our dataset collected from different scenarios. Based on the research insights, we discovered explainable physiological features and analysis methods that could be applied in both offline analysis and real-time analysis to generate biofeedback. The physiological dataset collected from in-the-wild group events (98 recordings from three dance performances and 48 recordings from three concert sessions) with offline analysis sample codes are freely available for the HCI community.

- Multiple real-time analysis algorithms embedded in wearable sensing and feedback systems and practical implications for biofeedback design deriving from the interdisciplinary collaborations. An exploration on detecting and sharing physiological experience, which opens up a potential direction to augment group dynamics by affecting and manipulating physiological experience.

1.5. Thesis Structure

The remainder of this thesis is structured as follows. Chapter 2 presents a literature review regarding physiological data as social signals, sensing live audience, entrainment, and live feedback in group events. Chapter 3 presents the framework for physiological data analysis. The framework is first explained by introducing the offline analysis component and the real-time analysis component. It further details three stages and main steps to analyze physiological data for understanding and augmenting group dynamics.

Chapter 4, 5, Chapter 6 describe the application and evaluation of the analysis framework on several studies (see the overview of three chapters in Figure 1.2 and related publications in Table 1.1). Chapter 4 describes studies where we explored analysis methods and features that could be used for quantifying and sharing inner feelings at group level. Chapter 5 presents further investigation on how to link collective physiological data to in-the-wild group experience. Chapter 6 describes the development of real-time physiological event detection algorithm that has been embedded in real-life biofeedback to influence the experience in group activities. Besides, it presents offline analysis on the collected physiological dataset from audience group to investigate the effect on group dynamics brought by biofeedback.

Chapter 7 presents discussions from the aspect of data and implications for applying the analysis framework in practice-led research. Finally, Chapter 8 concludes the thesis by reviewing research questions, reflecting on the limitations, and providing future prospects.

	Exploration (Chapter 3)		Linking (Chapter 4)	Influence (Chapter 5)
Scenario	Social Game	Online Learning	Contemporary Dance	Piano Concert
Dataset				
Small group (<=10)	✓	✓		
Large group (>10)			✓	✓
Multimodal	✓		✓	✓
Group Interaction				
In–person	✓		✓	✓
Online	✓	✓		
Under Facilitation	✓	✓		
Explicit	✓	✓		
Implicit			✓	✓
Analysis				
Offline Analysis	✓	✓	✓	✓
Real–time Analysis		✓		✓
Trend	✓	✓	✓	✓
Aggregation	✓	✓	✓	✓
Statistics	✓	✓	✓	✓
Similarity	✓	✓		✓
Prediction				✓

Figure 1.2 The overview of three chapters describing applications and evaluations of the framework. Each chapter has a specific emphasis and features regarding dataset, group interaction, and analysis. Only the analysis methodologies described in this thesis were listed and checked in the analysis segment.

Table 1.1 Publications included in this thesis with the respective chapter.

Chapter	Publication
Chapter 4	**Jiawen Han**, Chi-Lan Yang, George Chernyshov, Zhuoqi Fu, Reiya Horii, et al. "Exploring Collective Physiology Sharing as Social Cues to Support Engagement in Online Learning." In 20th International Conference on Mobile and Ubiquitous Multimedia (2021) [38]
Chapter 5	1. **Jiawen Han**, George Chernyshov, Moe Sugawa, Dingding Zheng, Danny Hynds, et al. "Linking Audience Physiology to Choreography." ACM Transactions on Computer-Human Interaction (2021) [39] 2. Sugawa, Moe, Taichi Furukawa, George Chernyshov, Danny Hynds, **Jiawen Han**, et al. "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 (2021) [40] 3. Zhuoqi Fu, **Jiawen Han**, Dingding Zheng, Moe Sugawa, Taichi Furukawa, et al. "Boiling Mind-A Dataset of Physiological Signals during an Exploratory Dance Performance." In Augmented Humans Conference (2021) [41]
Chapter 6	1.Yan He, George Chernyshov, **Jiawen Han**, Dingding Zheng, Ragnar Thomsen, et al. "Frisson Waves: Exploring Automatic Detection, Triggering and Sharing of Aesthetic Chills in Music Performances." Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 6, no. 3 (2022) [42] 2.Yan He, George Chernyshov, Dingding Zheng, **Jiawen Han**, Ragnar Thomsen, et al. "Frisson Waves: Sharing Frisson to Create Collective Empathetic Experiences for Music Performances." In SIGGRAPH Asia 2021 Emerging Technologies (2021) [43]

Chapter 2
Literature Review

This Chapter will introduce the premises that this thesis is based on and the context that the framework is developed. Section 2.1 presents the background knowledge from the social neuroscience field and explains how physiological data relates to social processes. Section 2.2 summarizes the recent works that applied sensing technologies to scenarios where a group of people attended. Section 2.3 discusses the concept of entrainment informing why this concept could be borrowed to develop the framework. Section 2.4 presents recent works where live feedback has been implemented in group events such as lectures and performances.

2.1. Physiological Data as Social Signals

The emotions that humans experience while interacting with their environment are associated with varying degrees of physiological arousal where ANS plays a crucial role [44, 45]. Emotional states associated with ANS responses can be inferred using physiological data like Electrocardiography (ECG), EEG, EDA, and Blood Volume Pulse (BVP) [46, 47]. ANS is mediated by two branches, which are the sympathetic nervous system (SNS) diverting energy outwards towards rapid mobilisation and environmental engagement during the fight-flight and the parasympathetic nervous system (PSNS) directing energy inwards towards processes of recuperation and self-care during the complimentary rest-digest [22, 48]. In my work, I mostly focused on BVP and EDA considering its unobtrusiveness and ubiquity enabling the potential to investigate group dynamics beyond individual sensing.

BVP is a pulse-based method of calculating the cardiac cycle from which the interbeat interval (IBI) can be inferred [47]. HR and heart rate HRV can be calculated based on IBI and are considered to result from the ANS activities. The neurovisceral integration model describes HRV as the result of prefrontal cortex activities that affect modulation of the PSNS and SNS nervous systems balance [49]. Hence, HR and HRV has been shown to be an indicator for reflecting emotions and a powerful tool for observing interactions between subjectivity and physiology either activated by PSNS (e.g. pNN50 [1]) or SNS (LF/HF ratio [2]) or both (SDNN[3]) [50–52].

EDA measures variations in skin conductance related to sweating and is a measure of the sympathetic nervous system. It is being used for over a century [53] and remains one of the most widespread tools for the measurement of autonomic nervous system responses in psychology and psychotherapy [54,55]. As a sensitive marker, EDA is often used to assess emotional arousal [56–59]. Two components of EDA could reveal different processes of EDA's time course. EDA tonic component indicates the slow change of skin conductance levels while the EDA phasic component reflects the quick and prompt change of skin conductance response [57,60,61].

Considering the correlation between affective states and physiological data, physiological data could be externalized as social cues. Recent works in HCI field have extended the idea of sensing physiological data to interacting with physiological data via biofeedback [22]. Hook et al. proposed somaesthetic appreciation to design feedback interactions focusing on bodily experience such as breath [26]. Besides enhancing the behavior of introspection, biofeedback has been explored as a medium to increase connection and empathy in social interactions [23–25]. In this thesis, we would like to focus more on the potential application of biofeedback from the perspective of social interactions in collective experience rather than individual response.

Table 2.1 Recent work about sensing live groups. Collection methods are classified by referring to the "In-the-wild HCI research" definition by Benford et al. [27] (m = *minutes*). Boiling Mind project (Chapter 5) and Frisson Waves project (Chapter 6) are two projects we evaluated the proposed analysis framework. HR/HRV/EDA data of large-scale audience groups were collected in-the-wild.

Measurement	**Scenario**	**Audience Numbers**	**Recording Duration**	**Sensing Technology**	**Collection Methods**
GSR/EDA	Dance performance video [5]	49	11 m	Thought Technology GSR fingerwraps	Lab
	Films in theater/festival [6]	34	130 m	Affective Q Sensor	In-the-wild
	Live performance [7]	15	28 m	Customized sensor	In-the-wild
HR/HRV	Dance performance [8]	24	63 m	Bioharness 3 Sensor	Lab
	Piano performance (live/recorded) [9]	37	70/50 m	Win Human Recorder	Lab
	Dance performance [10]	101	35 m	Empatica E4	Lab
BCI	Academic presentation [11]	11	35 m	Neurosky Mindwave	In-the-wild
Body Movement	Dance performance [12]	38	100 m	Night vision cameras	In-the-wild
	Music concert [13]	49	8 songs	Passive optical motion capture system	In-the-wild
	Dance/ talks/ music [14]	75	79/42/22 m	Customized neck-worn sensors	In-the-wild
EDA/HR/HRV	Boiling Mind Project [40]	98	3x70 m	Customized wrist band	In-the-wild
EDA/HR/HRV	Frisson Waves Project [43]	48	60 m	Customized wrist band	In-the-wild

2.2. Sensing Live Audience

Table 2.1 summarizes recent works that applied sensing technologies to explore group dynamics. Benford et al. [27] refer to the fusion of computing technology and public artistic projects (including performance) as “in-the-wild” research, in the sense of engaging “real” users with emerging technologies in real settings under demanding conditions of actual use, as opposed to the more constrained “lab” environment. We adopted this terminology and classify related work on audience sensing into either “lab” or “in-the-wild”, with the latter referring to recordings during actual live events. Compared to physiological methods, physical signals, like body movements, facial expressions, etc. are easier to record in-the-wild and thus feature prominently in the literature [12–14]. Theodorou et al. extracted face, hand and body movement data collected from four contemporary dance performances together with two follow-up surveys on selected audience members for ranking the performance and reporting engagement [12]. By comparing motion data with surveys’ results, they suggested lowest overall audience movement are perceived to be highest engagement but no systematic effect of dancers movements on audience movements. Gedik et al. developed an approach to predict audience self-reported binary experience (positive and negative) using accelerometer and proximity sensor data [14]. They also linked audience body movements to memorable moments that were reported. In live music contexts, head movements were faster during live concerts than album-playback concerts. While Swarbrick et al. explained this as higher engagement [13]. These differences in audience movement can be explained by the different performance types, and as the current study focuses on contemporary dance, we opt here to use physiological recordings instead of physical.

Previous work using EDA to track audiences includes Silveira et al.’s exploration of using viewer’s EDA to classify movie ratings [6]. Latulipe et al. used wearable EDA to record 49 participants watching a video of a dance performance. Their

1 Percentage of adjacent NN intervals that differ from each other by more than 50 ms

2 Ratio of low frequency (LF) to high frequency (HF) power

3 The standard deviation of the IBI of normal sinus beats

results show strong correlations between the EDA and self report data, which supports the validation of temporal EDA data as reflection of audience group's engagement [5]. However, since the audience only watched the recorded version of the performance, we could not ignore the difference of audience reactions between their study and those in real performance. Wang et al. recorded EDA from a live audience (15 participants for a 28-minute comedy) using wired electrodes on the palms [7]. From questionnaires' and EDA data, they clustered audience members and identified a strongly correlated main group. They uncovered events (e.g., "balloon pops") as changes in EDA and posited this as evidence of psycho-physiological engagement.

HR/HRV in the group has been mostly studied in lab settings. Shoda et al. conducted a series of experiments to explore how audience members' HR and the spectral features of HRV differ between music that is live versus recorded, and fast tempo versus slow tempo. They show that audiences tend to have higher HR and lower sympathovagal balance when listening to faster live pieces. The sharing interaction between pianists and the audience could reduce audience's physiological stress [9]. In Vicary et al.'s study, they tracked dancers' acceleration as movement data and the audience' HR as affective feedback over five live performances. Their results indicate that movement synchrony among performers could predict audience aesthetic appreciation [10]. Instead of looking into the synchrony among performers, Bachrach et al. used Myriam Gourfink's choreography design to modulate respiratory rate and internal temporal clock and investigated the entrainment of audiences and dancers during dance performances. They carefully designed four experimental sessions from which they collected respiratory rate, and questionnaires related to subjective engagement and time perception [8]. Their work suggests that attention to breathing is closely related to entertainment. Those previous studies inspire us to find connections between physiological data and certain aspects of group dynamics such as engagement and entrainment.

2.3. Entrainment

Behavioral entrainment has been firstly explored since 1960s based on video analysis of movement [62]. Condon et al. examined the movement regularities of people in communicative contexts by coding the trajectory of visible movements of participants' body parts. Their follow-up research discovered the "bond" between human beings as an expression of participation within shared organizational forms rather than as isolated entities [19]. Further advanced in motion tracking methods such as attaching accelerometers to the interacting individuals revealed the possibilities of investigating group entrainment multidimensional and continuous sensing data [63,64]. Lang et al. measured the acceleration of hand movements of participants while hearing three different auditory stimuli and discovered the exposure to musical rhythm enhanced behavior coupling [64]. Previous works have proved entrainment in social interactions could be indicative for group cohesion, psychological connectedness, and inter-subjective engagement investigated through not only behavioral [65,66] but also physiological measurements [8,37,67,68]. Indexed by continuous measures of the ANS, interpersonal autonomic physiology (IAP) describes the relationship between people's physiological dynamics and one of the common observations is the interdependence or association in partners' physiological activities [4]. Multiple terminologies (e.g.physiological linkage, physiological synchrony, physiological coherence) have been used to describe the phenomenon and we adopted the "physiological entrainment" in this thesis.

The term "entrainment" refers to the process by which independent rhythmical systems interact with each other so that they adjust themselves and eventually become rhythmically coupled [8, 18]. The notion was borrowed from physics with the classic mechanical example of pendulum clocks [17]. In social contexts, entrainment is mostly related to behavioral evidence that enables co-acting individuals to perceive and produce rhythmic movement via perception-action link [69]. Different from behavioral measurements such as gestural expressions, physiological entrainment is less controllable and noticeable. However, as a component of shared experiences, physiological entrainment could be used as an objective measure of internal processes accompanying empathic interactions [37] and could further indicate group cohesion, connectedness, and engagement in group dynamics [8, 37, 67, 68]. Therefore, the physiological data analysis framework I proposed centers around the concept of entrainment: What metrics calculated from physiological data could be used to reflect entrainment? Can we know when does the entrainment happen through the physiological data patterns? How can we apply the findings to real-time biofeedback practices?

The exploration of entrainment in this thesis mainly focuses on the similarity between timeseries data of physiological features as entrainment has been proved to be related to the presence of similar reactions among group members [37]. Various analysis methods have been explored to quantify entrainment in previous works such as dynamic time warping (DTW) [67, 70–72], pearson correlation [73, 74], cross recurrence quantification analysis (CRQA) [37], wavelet coherence analysis [68, 75], and machine learning algorithms [76, 77]. We selected DTW as our measurement in this thesis. One reason is that DTW could be applied to two time series data with different lengths and thus would be flexible to use. Another reason is that DTW is relatively easy to interpret the results, which could be a suitable start for HCI researchers and practitioners to explore physiological entrainment.

Moreover, there are some analysis methods that could help researchers understand the entrainment experience besides directly calculating the similarity between physiological data. Statistical analysis (e.g. regression analysis and descriptive statistics) [78] could show the overall trend and fluctuations in the group dynamics. With an adequate amount of data, classification and regression through machine learning methods could be applied. Clustering has also been proved as a potential analysis method to understand subgroups' experiences [7].

2.4. Live Feedback in Group Events

With wearable sensing devices, live feedback could be implemented in real-time to create novel interactions in group events. Hassib et al. presented a system to infer and visualize the audience's implicit engagement from brain waves [11]. The real-time view shows the current audience's average normalized engagement score at that time and tracks the dynamic changes over time. There are more works exploring live feedback in the field of performance as a valuable test bed for integrating technologies with real-life situations [27]. One direction is to create connections between performers and stage design. Rodrigues et al. adopted Kinect tools to develop a system where projections on the stage could respond to dancers' movement [79]. Brown et al. introduced motion capturing to allow the body movements of dancers to drive real-time music generation and arrangement [80]. Another direction is to create live feedback based on the audiences' senses and feelings and integrate it into the performance environment. Lindinger et al. invited the audience to co-create the visual performance environment by sending text messages to create the character clouds. Performers interacted with the character clouds through a Kinect body-tracking system [81]. Rostami et al. [82] explored how to use physiological sensing and bodily tracking technologies for artists to engage the audience through two design workshops. Moreover, Khut et al. [22], Hook et al. [83], and Benford et al [27] explored design possibilities to create various types of biofeedback including visual, sonic, ambient, and haptic interactions. Their explorations manifest several core concepts (e.g. biofeedback design and affective loop) and frameworks guiding in-the-wild research.

2.5. Summary

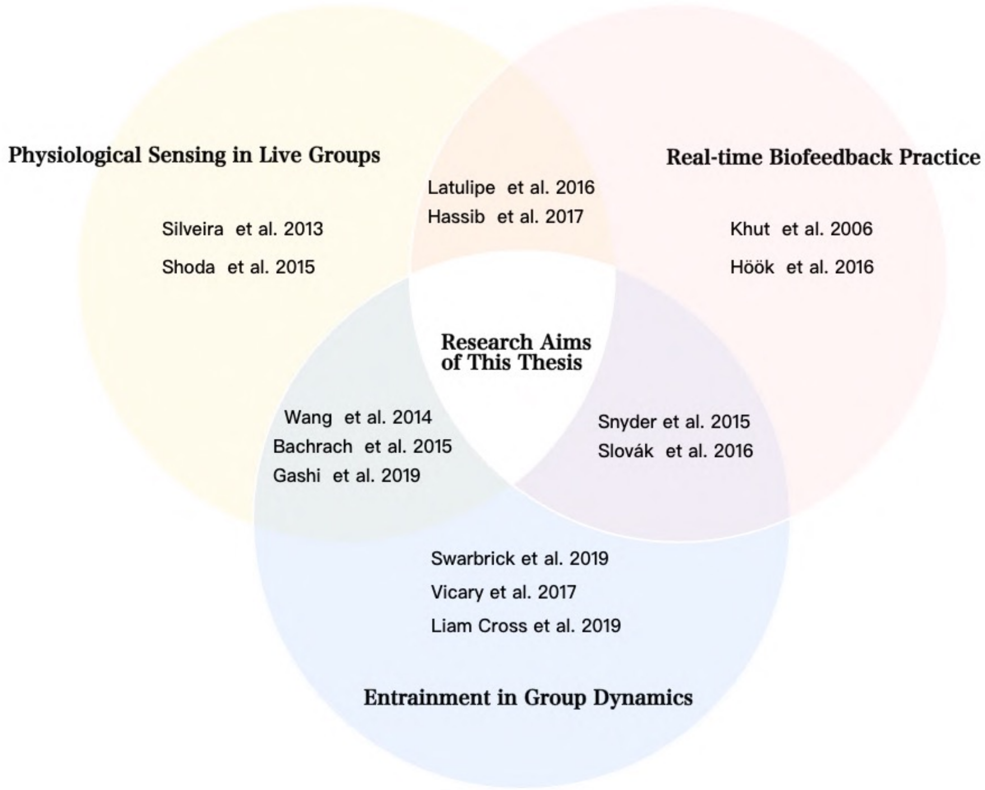

Figure 2.1 Research aims of this thesis in the scope of related works. Significant works in the related fields: physiological sensing in live groups, real-time biofeedback practice, and entrainment in group dynamics.

A number of previous works have investigated entrainment in group activities from behaviors and movements [10, 13, 66]. Swarbrick et al. recorded the head movement data from participants attending live and album-playback concerts. They defined the frequency of participants' head movement to the beat of the song as the degree of entrainment and found that self-reported fans exhibited higher entrainment [13]. Vicary et al. proved movement entrainment among performers could predict aesthetic appreciation [10]. Cross et al. presented a comprehensive discussion about the prosocial effects of entrainment in group processes such as the increasing sense of belonging [66]. Integrating this concept with physiological sensing, researchers have explored physiological entrainment as a measurement to quantify collective experience such as enjoyment and engagement [7, 8, 67]. The entrainment of breathing rates was calculated and suggested a close relationship with attention to breathing in Bachrach et al.'s work. Wang et al. found similar EDA patterns occurred among participants with similar self-reported enjoyment and "cheerful" experience [7]. Gashi et al. applied DTW analysis to EDA data collected from an audience group and suggested physiological entrainment measured using DTW could be adopted as a proxy to quantify engagement [67]. Referring to the previous works, the research described in this thesis further explores practical ways of analyzing physiological entrainment to understand group dynamics during in-the-wild activities.

Moreover, physiological sensing has been widely adopted to generate novel interactions via biofeedback [22, 26]. The idea of sharing internal experience to create the feeling of being connected presents biofeedback's potential to elicit empathy and entrainment [23, 25]. MoodLight developed by Snyder et al. is an ambient lighting system that could adjust the colors of the light to users' arousal level suggested by EDA data [25]. The ambient feedback was tested also during paired interaction as an additional social cue to represent entrainment. Slovak et al.explored biofeedback via heart rate data and found the potential of sharing heart rate to support connectedness in social interactions [23]. However, the scenarios are usually based on dyadic interactions and interpersonal relationships. This thesis extends this practice to biofeedback implementation in group events and explores how to use entrainment to augment group interactions. To approach the aims, this thesis will follow the framework for physiological data analysis described in Chapter 3.

Chapter 3

Framework for Physiological Data Analysis

This work introduces a practical framework for analyzing physiological data collected from group events. Ubiquitous computing and wearable sensing have provided opportunities for researchers and practitioners to make use of sensor data to understand human behavior and design sensor-based interactions [36, 84]. Compared with physiological data, physical data (e.g. OpenPose [85]) has been more explored and embedded in systems for gesture estimation. Most sensor data analysis software packages either target professional engineers with a high threshold for use (e.g. LabView [1]) or lack of the availability to transform sensor data into features and high-level concepts(e.g. Exemplar [36]). The proposed framework aims to support HCI researchers and practitioners to process, analyze, and interpret physiological data. The concept of entrainment will be adopted to understand and augment group dynamics in terms of analysis methodologies and interpretation perspectives in the two components: offline analysis and real-time analysis.

3.1. Components of the Proposed Framework

Figure 3.1 summarizes the framework for physiological data analysis by explaining the analysis flow of two components separately: offline analysis and real-time analysis.

Offline analysis is conducted on the recorded datasets to support reflecting group dynamics during past group interactions. With less concern about computational complexity, offline analysis allows statistical calculations and even more

1 `http://www.ni.com/labview`

advanced time-series data analysis methods. The aspects to interpret the data results and findings could be more flexible and comprehensive.

On the contrary, real-time analysis is to support not only researchers but also practitioners and participants to make meaningful inferences on the flow and events during the group interactions equipped with biofeedback systems. Real-time in computing processes should guarantee a relatively short response time to trigger feedback. Therefore, the computation time and memory storage required to execute the analysis should be carefully considered in each stage of data analysis.

3.2. Stages of Data Analysis

Based on established data analysis and signal processing procedures, five main steps are summarized [86, 87]: data collection, preprocessing, feature extraction, analysis, and interpretation. Figure 3.1 illustrates the analysis flow by grouping five main steps into three stages. The remaining contents of this section describe the procedures over three stages. By applying the proposed framework to four projects (Chapter 4, Chapter 5, and Chapter 6), we further explore and summarize how to link the physiological data results to group dynamics in the discussion chapter (Chapter 7) especially focusing on 1) selecting and extracting appropriate and explainable features, 2) analyzing and aggregating individual data to reflect collective experience at group level, 3) interpreting and discussing the findings centering around the concept of entrainment in group dynamics.

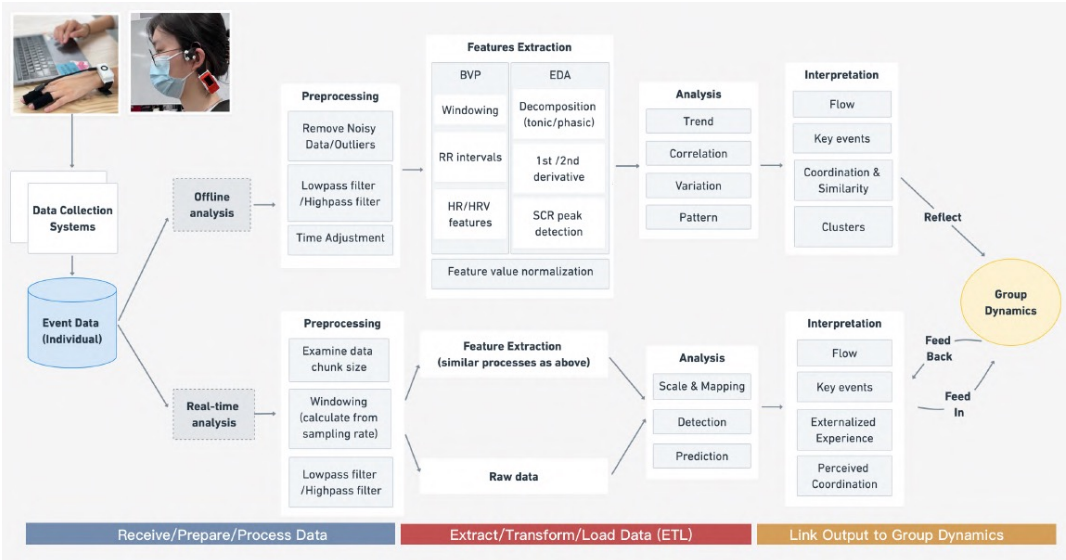

Figure 3.1 Proposed physiological data analysis framework to understand and augment group dynamics.

3.2.1 First Stage: Receive, Prepare, and Process

Data collection

Large scale physiological data could be collected by either commercial sensing devices or customized ones (see Table 2.1 for the summary of sensing devices used in recent research works). Section 3.3 describes the devices used for collecting BVP and EDA data used in the projects mentioned in this thesis. There are also various types of physiological responses that could be investigated from physiological data such as brain activity from EEG data, cardiac activity from BVP data, and electrodermal activity from EDA data. We mainly explored BVP data and EDA data in this thesis because of the following reasons. Firstly, BVP and EDA data could be collected through devices worn on the fingers and wrists, which is less complicated to set up a group of participants. Another reason is that physiological responses revealed by BVP and EDA data (e.g. heart rate acceleration and more sweating) are relatively familiar to most people. However, other types of physiological data could also be analyzed following the steps in the proposed framework by adjusting preprocessing parameters and extracting appropriate features.

For offline analysis, data could be saved locally. The analysis could be conducted via python or other coding languages at existing programming platforms such as Jupyter Notebook[2]. For real-time analysis, data could be streamed to servers via User Datagram Protocol (UDP) and processed in code scripts (python script in the proposed framework) via command line tools or integrated development environment (IDE) such as Pycharm[3].

Preprocessing

The first step in preprocessing is to clean noisy data, missing data, and outliers by replacing the data points with zero, interpolation based on existing data points and removing the period of data where intense movement is detected. Secondly, as an essential step in signal processing, filters (e.g.lowpass, highpass, and bandpass filters) need to be applied to smooth and resample the signals. The parameters of the filters (e.g.order and cutoff frequency) could be adjusted referring to the normal range of signal frequency proved by related works in the psychophysiology field. Table 7.1 summarizes the information about the sampling rate and filter parameter adopted in the analysis described in this thesis.

3.2.2 Second Stage: Extract, Transform, and Load

Feature extraction

From BVP data, interbeat interval (IBI) could be inferred and used to calculate HRV features reflecting the variance of heart beat activity. HRV features are usually extracted within a subset of data points in the time series, which is known as sliding window (or rolling window). The window consecutively rolls back, holding the same number of data points within the window as it moves along the time series data stream. For offline analysis, cutting sliding window could be conducted after timestamp adjustment. For real-time analysis, considering the running time cost, instead of cutting sliding window, the chunk size of data streamed could be

2 `https://jupyter.org/`

3 `https://www.jetbrains.com/pycharm/`

adjusted in advance. HRV features could reflect physiological responses activated by either PSNS or SNS or both [52]. The onset of strong emotions is typically characterized by noticeably increased sweating on the skin. Thus, we mainly investigated the drastic changes in EDA data or in specific components of EDA (tonic and phasic) [57, 60, 61]. Section 3.3 summarizes and explains the features extracted in the analysis described in this thesis.

Analysis

Statistical analysis (e.g. regression analysis and descriptive statistics) could be adopted to observe the changes in group dynamics. Common descriptive analysis (e.g. calculating mean and median values) could also be used as real-time analysis to visualize data in the form of charts and graphs [11]. Predictive analysis could be applied with an adequate amount of data through machine learning methods to recognize and detect affective experience [47, 88]. Exploratory data analysis (e.g. clustering) could be used to explore the unknown correlations between individuals' experiences [7]. To further explore the concept of entrainment in the group, quantifying similarities between group members' physiological data and triggering similar physiological experiences are considered as the main focus of the analysis in this thesis. As summarized in Section 2.3, various analysis methods, such as DTW [67, 70–72], pearson correlation [73, 74],CRQA [37], wavelet coherence analysis [68, 75], and machine learning algorithms [76, 77],could be adopted to investigate entrainment. In this thesis, we explored multiple analysis methods including comparing trends and variance, calculating similarity by DTW, and detecting physiological events by machine learning algorithms. We selected the above analysis methods considering the difficulty of implementing and interpreting the algorithms in both offline and real-time analysis. However, other mentioned analysis methods are also worth further exploration.

3.2.3 Third Stage: Link the Outputs to Group Dynamics

Interpretation

To interpret the results revealed in the physiological data, we could start by understanding which ANS branch is related to the physiological feature (e.g. pNN50 is closely related to PSNS and higher pNN50 may suggest increasing relaxation). Subjective feedback collected by questionnaire or interview could provide proof to interpret the results when comparing sub-groups' reactions or labels to develop supervised machine learning models. Moreover, the development of group events (e.g. storyline and key moments) could be annotated to understand group dynamics or trigger biofeedback.

3.3. Devices and Features in this Thesis

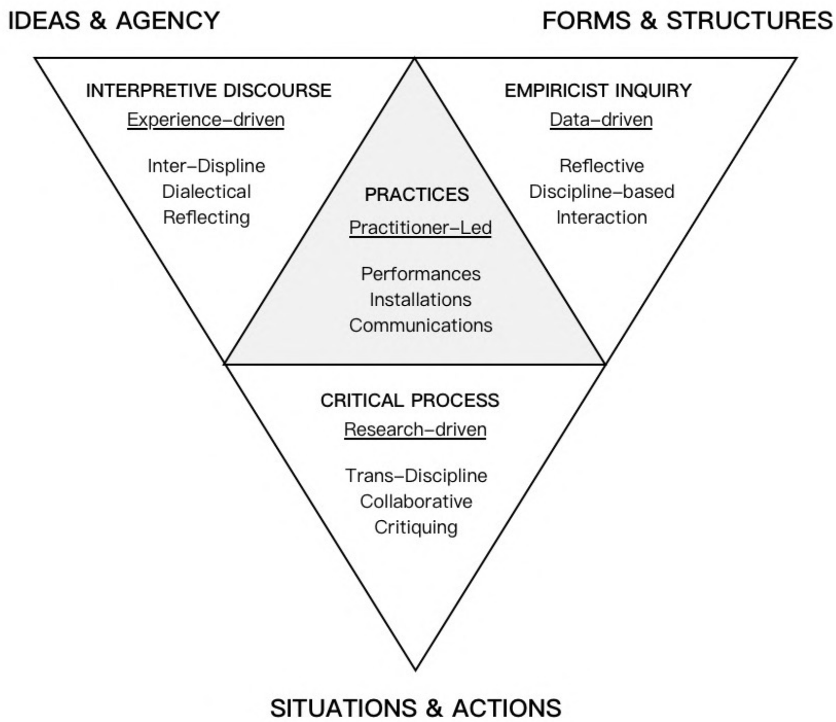

Figure 3.2 Practice-based methodology centers around practitioner-led practices. Sullivan et al. developed a framework explaining how to use art practice as research, which covers four primary aspects: interpretive discourse, empiricist inquiry, critical process, and practices [28]. Referring to Khut et al.'s theory for Biofeedback Artworks [22] and Benford et al.'s theory for Performance-led Research in-the-wild [27], we selected and added details based on our interdisciplinary collaborations to understand and augment live group experience with physiological data.

Chapter 4, 5, and 6 describe how we applied the proposed analysis framework to understand and augment group dynamics in various group interactions. Figure 3.2 presents a framework initially developed by Sullivan et al. which provides a guide for our interdisciplinary collaborations. Centering the aim of "practices", sensing devices, biofeedback design, and analysis methodologies were explored during iterations. This section describes the customized devices used for data collection and features generated for real-time and offline analysis.

3.3.1 Devices

In this dissertation, we selected EDA and BVP data for analysis. Figure 3.3 presents three major devices we used for data collection and recording.

Figure 3.3 Sensing devices used for data collection and recording in this dissertation. (a) and (b) were wrist-worn sensing devices while (c) was ear-based device.

Wrist-worn devices (Figure 3.3(a) and (b)) were principally developed by Gheorghe Cernisov (hardware and software development to save and stream physiological data) [89] and Dingding Zheng (sensor modality selection and platform design based on psycho-physiological knowledge) [90]. Yulan Ju designed the appearance of the wristband case shown in Figure 3.3(b). The device in Figure 3.3(a) was used in the project described in Chapter 5 to collect and stream EDA, BVP, and accelerometer data. The device in Figure 3.3(b) was used to collect and stream EDA and BVP in the projects described in Chapter 4 (Section 4.2 and Section 4.3), and Chapter 6.

Ear-based device (Figure 3.3(c)) was principally developed by Kanyu Chen [91] (hardware development and ear-phone prototype design) and Ziyue Wang (data streaming system). The device in Figure 3.3(c) was used to collect and stream BVP data in the project described in Chapter 4 (Section 4.4).

3.3.2 Features

This section summarizes HRV and EDA features we investigated in this thesis. Except features with generally accepted terminologies (e.g., pNN50 and SCR peaks), some features (e.g., EDA difference and EDA extrema) were given interpretive names to avoid ambiguity. Appendix C presents examples of extracting HRV and EDA features in python.

From BVP data, interbeat interval (IBI) is calculated. After removing abnormal beats from IBI, we got the IBI of normal sinus beats, which is usually referred as NN intervals. HRV features are generally calculated from NN intervals. For example, the average of NN intervals within a certain period is **Mean NN**. From the Mean NN, we calculated **Beats per Minute (BPM)** that usually refers to average HR.

HRV features are either related to SNS or PSNS activation or both [52]. In this thesis, we extracted pNN50, RMSSD and LF/HF ratio for our analysis. **pNN50** refers to the percentage of adjacent NN intervals that differ from each other by more than 50 ms. **RMSSD** refers to the root mean square of successive differences between normal heartbeats. Both of pNN50 and RMSSD are time-domain features that could quantify the amount of HRV observed during monitoring periods [52]. As closely related to the PSNS activation, pNN50 and RMSSD are usually negatively related to increasing arousal and corresponding affective states. Specifically, increasing pNN50 could reflect relaxation, engagement in the controlled process, and sustained attention [47, 92, 93]. RMSSD has been proved to be negatively correlated to stress and cognitive load [94, 95]. **LF/HF ratio** refers to the ratio of low frequency (LF) to high frequency (HF) power and could be an indicator of the balance between SNS and PSNS activity [52]. Even though the interpretation of the LF/HF ratio is controversial [96], it is still possible to explain the changes with cautious consideration of the recording contexts [97]. In most of cases, increasing LF/HF ratio could be related to rising arousal implying anxiety and excitement [88, 98, 99].

EDA is used to define autonomic changes in the electrical properties of the skin and includes tonic and phasic components [87]. Some previous works have used skin conductance level (SCL) and skin conductance response (SCR) to name tonic and phasic components. EDA tonic component indicates the slow change of skin conductance levels while the EDA Phasic reflects the quick and prompt change of skin conductance response [57, 60, 61]. In this thesis, we adopted **EDA Tonic** as the tonic component value and **EDA Phasic** as the phasic component value to directly distinguish the two features from their characteristics. However, we adopted **SCR peaks** to describe the peaks in EDA Phasic because this term has been generally accepted in python packages for signal processing (e.g. Neurokit2 [4] [100]). Besides established EDA features, we also explored **EDA difference** by calculating the changes in EDA response (the first derivative of the EDA data) and **EDA extrema** by looking specifically into the timings when EDA data drastically increased. For features derived from peak detection, we aggregate individual features into collective ones by counting the number of people who have experience peak events within a certain time period such as **EDA extrema count**.

Above are the principal features generated in this research. Features used in the specific studies were listed in each chapter's overview and marked if they have been used in real-time analysis implemented in biofeedback systems.

4 `https://neuropsychology.github.io/NeuroKit/`

Chapter 4

Explorations of Physiological Data in Groups

4.1. Overview

This Chapter describes the explorations of physiological data as social signals to help us understand group dynamics and communicate within group interactions. With the outbreak of the pandemic, online communication has become an essential part of group interactions yet usually keeps group members feeling distant. Two projects (social game project described in Section 4.2 and online learning project described in Section 4.3 and Section 4.4) reported in this Chapter aim to investigate and enhance the social bond in online group interactions. The social game project is to compare players' physiological responses when they are playing in-person and online. The online learning project is to explore feasible methods to share physiological data as novel interactions during online lectures and online workshops. Both projects provided valuable datasets to initially explore explainable physiological features and analysis methods to reflect group dynamics.

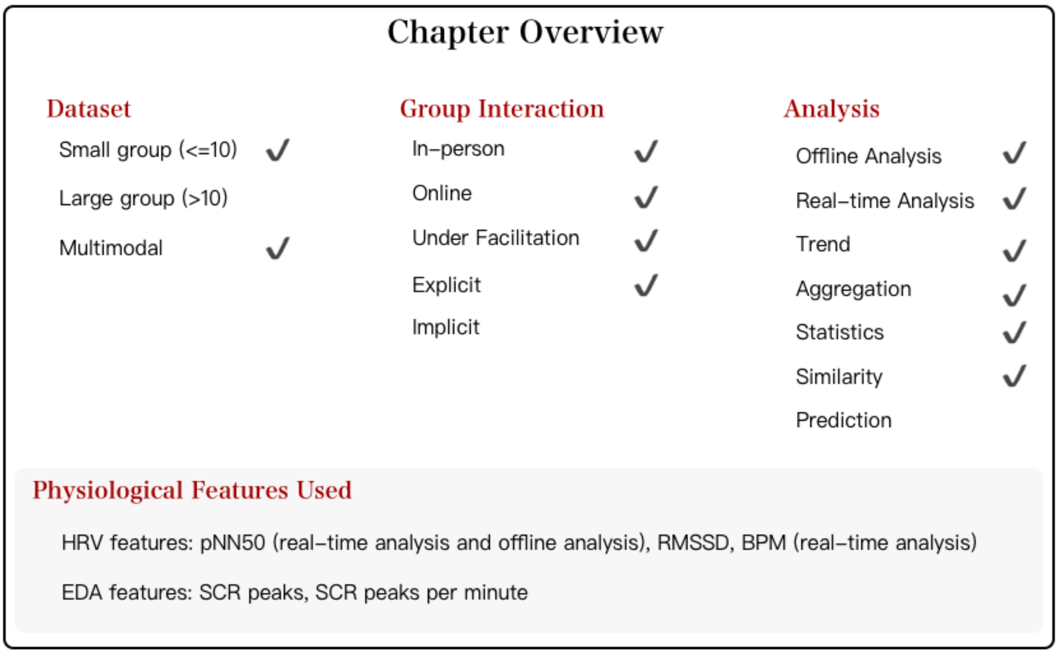

Figure 4.1 Overview of the project information in Chapter 4

Section 4.2 describes the project investigating group interactions during in-person and online werewolf games. In werewolf game, players are given specific roles and then belong to either good people side (villagers and gods) or bad people side (werewolves). As a distinguishing feature in werewolf game, concealing information and deception could be reflected in temporal changes of physiological data [101]. Werewolf game has been considered as a suitable activity to investigate group interactions because the game follows a relatively fixed structure and players need interact actively to understand the situation and collaborate within groups [102]. Moreover, we recorded physiological data when participants played in-person and online via Zoom, which enables a comparison and discussion. Therefore, this dataset could be a start point to initially explore physiological data from the perspectives of different groups (good people side and bad people side) and different interaction medium (in-person and online).

Section 4.3 and Section 4.4 describe the projects where we explore the real-time analysis implementation to share physiological data within groups and its potential effect. We selected online learning conditions for the setup. In many online learning cases, people have reported to feel less connected and engaged due to insufficient social cues [103–105]. Besides testifying the real-time analysis setup, we would also like to investigate whether sharing physiological data could enhance the feeling of being connected.

Major parts in this chapter (Section 4.3) are based on the following research paper we published:

- **Jiawen Han**, Chi-Lan Yang, George Chernyshov, Zhuoqi Fu, Reiya Horii, Takuji Narumi, and Kai Kunze. "Exploring Collective Physiology Sharing as Social Cues to Support Engagement in Online Learning." In 20th International Conference on Mobile and Ubiquitous Multimedia, pp. 192-194. 2021.

4.2. Quantify Group Dynamics

4.2.1 Data Collection and Datatset Description

We recorded nine players' EDA and BVP data when they were playing in-person (see Figure 4.2) and playing online (see Figure 4.3). Nine players were all native Chinese speakers (self-identified as female=5; male=4) aging between 22 and 31 years (mean = 26.3, SD = 2.57) and were familiar with the werewolf game rules. However, one player's data were excluded due to internet failure. BVP and EDA data recorded were with different length due to the different time duration of each round of game. We explained the study and the game rule in case and received players' consent before the game started. We firstly let players play in-person and then invite them to be physically apart and joined the game via Zoom. Players' identities were not controlled and assigned randomly. The identities we selected in the game were as follows:

- Villagers: Three villagers. Close eyes during the night stage and do not know the members in the same group.
- Gods: One seer who could know one player's identity during the night stage (either good or bad identity). One witch who could either heal or poison

one player during the night stage. Hunter who can “kill” one player after he/she is “killed” or voted out.

- Werewolves: Three werewolves. Know the other group members and “kill” one player during the night stage after slightly communication with each other.

Figure 4.2 Photo taken when participants played the werewolf game in-person.

Figure 4.3 Screen shot when participants played the werewolf game online via Zoom (Mosaic was applied to the picture for privacy concerns).

4.2.2 Analysis Process

BVP

A 2^{nd} order Butterworth low pass filter (from python package, *scipy.signal*) [1] was then used to cut high frequency noise above 3 Hz [106, 107]. HRV features were calculated every four minutes with a two-minute sliding window. For feature selection, we would like to choose a relatively stable feature that could reflect the beat-to-beat variance in heart rate. Therefore, we choose **RMSSD**[2] that has been proved to be negatively correlated to stress and cognitive load [94, 95]. Following the concept of entrainment, we firstly investigated the trends of group dynamics during each round of the game. Since there are good people side and werewolf side,

---

1 https://scipy.org/

2 The root mean square of successive differences between normal heartbeats

HRV features were averaged within each side and over all the players. Figure 4.4 shows the overall trend and the trends for each side. We further looked into the similarity and correlation between each pair of players' RMSSD during two rounds. Dynamic Time Warping (DTW) implemented with dtw-python module[3] was used to compare quantitatively [108]. The higher the value of distance calculated by DTW the lower the entrainment them between two timeseries data and vice-versa. We applied paired t-test to investigate whether significant difference existed between in-person game round and online game round in terms of physiological entrainment.

EDA

A 2nd order Butterworth low pass filter (from python package, *scipy.signal*) was then used to cut high frequency noise above 0.5 Hz [106, 107]. We extracted EDA features by focusing on the changes in EDA responses especially the peaks in the phasic changes as proved to be related to sudden aroused feelings [57] – known as **SCR peaks**[4]. We counted the number of players who had experienced SCR peaks every two minutes to represent the collective trend of arousal dynamics as aggregated SCR peaks (Figure 4.5). We further looked into the similarity and correlation between each pair of players' SCR peaks per minute during two rounds. DTW implemented with dtw-python module was used to compare quantitatively [108]. The higher the value of distance calculated by DTW the lower the entrainment them between two timeseries data and vice-versa. We applied paired t-test to investigate whether significant difference existed between in-person game round and online game round in terms of physiological entrainment.

4.2.3 Results

The trends of group dynamics are reflected in both HRV (RMSSD) and EDA features (aggregated SCR peaks). Compared with In-person round, all the players experienced more fluctuated RMSSD while playing the game online. The key

3 https://dynamictimewarping.github.io/python/

4 Peaks of skin conductance response or phasic components of EDA (EDA Phasic in this thesis)

events for the werewolf side to execute "kill" actions are highlighted in yellow. Figure 4.4 shows the trends of RMSSD both over all the players and two sides. During In-person round, the average RMSSD value among werewolf side is higher than the other side. While during Online round, the average RMSSD value among werewolf side is lower than the other side.

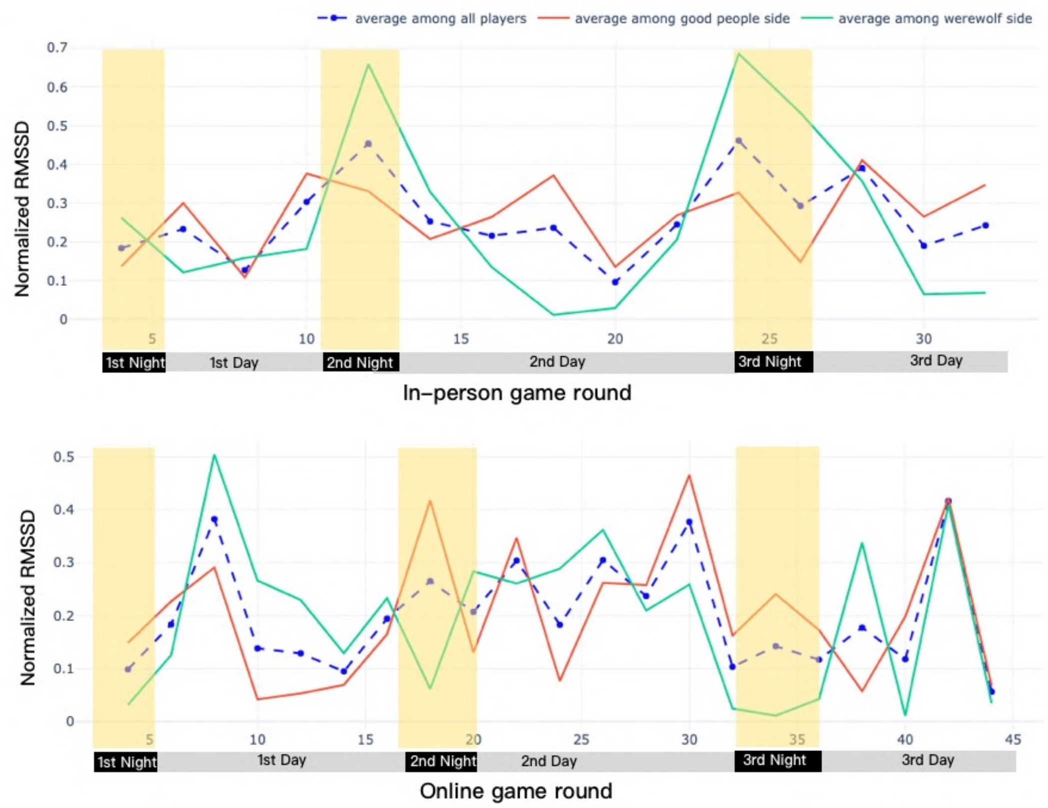

Figure 4.4 The average RMSSD values during the in-person (top) round and online round (bottom). The green line shows the average RMSSD value among werewolf players. The red line shows the average RMSSD value among the good people side. The Blue dashed line shows the average RMSSD value among all the players. Key events where werewolf players execute "kill" actions are highlighted in yellow.

The overall aggregated SCR peaks during in-person round (mean = 12.75, sd

= 1.95) are obviously higher than those during online round (mean = 2.74, sd = 1.51). The maximum aggregated SCR peaks exist either during or before the "Night" comes in both in-person and online rounds (see Figure 4.5).

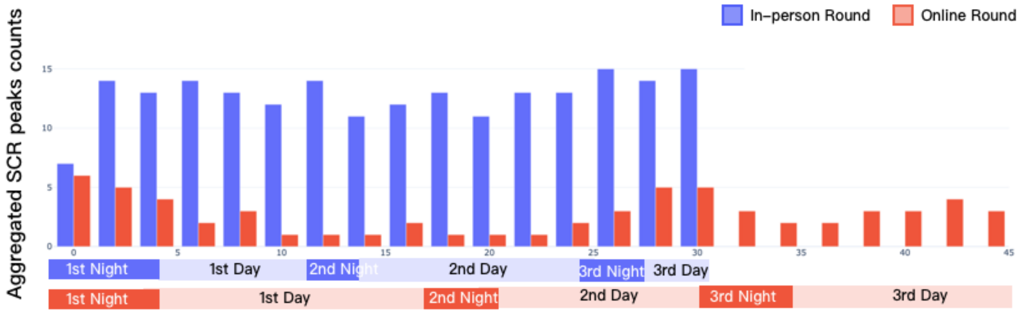

Figure 4.5 Aggregated SCR peaks occurred during in-person round (blue color) and online round (red color). The timelines of each game round were illustrated below the bar chart in corresponding colors.

For intersubject analysis, normalized distance calculated by DTW was adopted to quantify the similarity between every two players' timeseries data (RMSSD and SCR peaks per minute).

According to the paired t-test, normalized distance of RMSSD in In-person round (mean= 0.14, sd= 0.036) is significantly lower than that in Online round (mean = 0.18, sd=0.071) ($t(27)= -2.23, p<.05$). Figure 4.6 shows the overall distribution of normalized distance for each pair's RMSSD timeseries data.

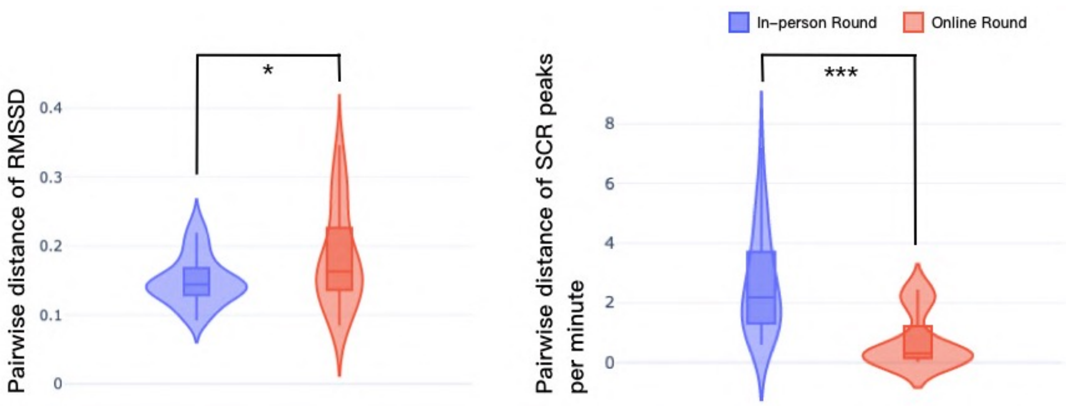

Figure 4.6 Comparison of paired distance of RMSSD (Left) and SCR peaks per minute (Right) between every two players during in-person round and online round. The inner box plot shows the show the minimum, first quartile, median, third quartile, and maximum values of timeseries normalized distances of pairwise physiological data's distance. The smaller the distance, the more similar the pair of timeseries data. The outer smoothed violin shape illustrates probability density. The width of the shape indicates how frequently certain values occur. ($*p<.05, ***p<.001$)

According to the paired t-test, normalized distance of SCR peaks per minute in In-person round (mean = 2.66, sd= 1.708) is significantly higher than that in Online round (mean = 0.72, sd= 0.887) ($t(27)= 4.77, p<.001$). Figure 4.6 shows the overall distribution of normalized distance for each pair's timeseries data of SCR peaks per minute.

4.2.4 Subjective Feedback

Methodology

To understand the group dynamics during in-person and online game, we asked nine players to fill out surveys measuring sense of community, emotional engagement, psychological engagement, and sense of co-presence in 7-Likert scale [105, 109, 110] after each round.

Questions to measure sense of community were as follows (Likert scale: "1-Strongly Disagree" to "7-Strongly Agree") :

- I felt like a member of this game.
- I did not belong in this game.
- I felt connected in this game.
- I felt that I matter to other players in this game.
- I had good bond with other players or teammates in this game.
- I felt distant from other players.

Questions to measure emotional engagement were as follows (Likert scale: "1-Strongly Disagree" to "7-Strongly Agree") :

- I have done my job well in the game.
- It is easy for me to understand other players in the game.
- I do not have friends in this game.
- We have a nice team spirit in the game.

Questions to measure psychological engagement were as follows (Likert scale: "1-Strongly Disagree" to "7-Strongly Agree") :

- I paid close attention to other players.
- I was easily distracted from other players when other things were going on.
- Other players paid close attention to me.
- I tended to ignore other players.

Questions to measure sense of co-presence were as follows (Likert scale: "1-Strongly Disagree" to "7-Strongly Agree") :

- I often felt as if other players and I were in the same environment together.
- I think other players often felt as if we were in the same environment.

Results

The analysis of the survey results was mainly to investigate whether players experienced different group dynamics when they were playing in-person and online, and whether the implementation of biofeedback could augment the group interaction.

We applied paired T-test between the survey answers from In-person no biofeedback round and Online no biofeedback round. Sense of co-presence during In-person round (mean = 5.89, sd= 0.89) is proved to be significantly higher than that during online round (mean = 4.64, sd= 0.91), $t(8) = 3.75$, $p<.005$. We did not find statistically significant difference in terms of sense of community ($p = .19$), emotional engagement ($p = .42$), and psychological engagement ($p = .20$) (see more detailed information in Figure 4.7).

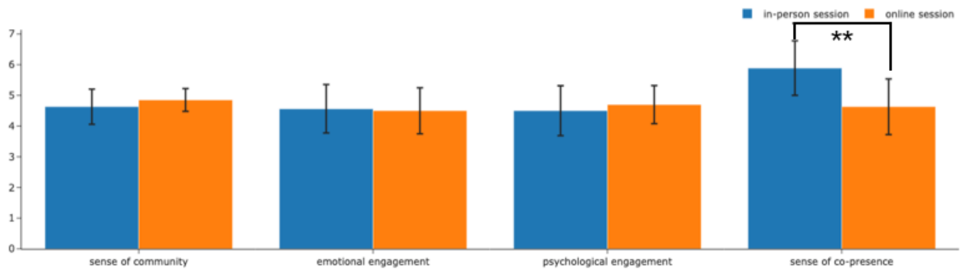

Figure 4.7 Summary of the survey results including sense of community ((in-person: mean = 4.63, sd = 0.57; online: mean = 4.85, sd = 0.37)), emotional engagement (in-person: mean = 4.56, sd = 0.79; online: mean = 4.5, sd = 0.75), psychological engagement (in-person: mean = 4.50, sd = 0.81; online: mean = 4.70, sd = 0.62), and sense of co-presence (in-person: mean = 5.89, sd = 0.89; online: mean = 4.63, sd = 0.91).

4.2.5 Discussion and Interpretation

Trends of Group Dynamics

Figure 4.4 shows the trends of RMSSD during both in-person round and online round. Based on visual inspection, we could notice more fluctuated lines occurred during online round, which may suggested a more unstable state in terms of cognitive load [94, 95]. Figure 4.5 shows the trends of aggregated SCR peaks in the form of bar chart. As related to sudden arousal increase indicating excitement and anxiety, SCR peaks occurred more during the in-person game [57]. One reason could be playing in-person could enable more observation on unconscious non-verbal behaviors thus increase the anxious feelings. However, due to the ordering effect, players started the in-person game first and might feel anxious in the beginning and get ease to the game when time passed.

Werewolf game follows a determined structure consisted of Day (players discuss and vote for one werewolf player) and Night (werewolves decide to "kill" one player while the others close their eyes). Therefore, Night could be considered as key events because werewolves need to discuss and decide in very short times while the others are usually anxious to wait for the coming results. Obvious fluctuations in RMSSD values could be observed in both in-person and online game rounds (see the highlighted regions in Figure 4.4). However, RMSSD reached peaks when Night came during in-person round while showed opposite tendency during online round. This result may suggest the different mental states for werewolves exist when they were playing in-person and online. However, players who were werewolves were not same, it is hard to draw a clear conclusion. Aggregated SCR peaks in Figure 4.5 reached high values around the key moments, either before the Night came (the time when players voted for a werewolf) or during the Night.

Physiological Entrainment among Group Members

The results calculated from DTW shows the physiological entrainment during the game. Players during the in-person round present significantly more entrained in terms of RMSSD while significantly less entrained in terms of SCR peaks per minute. Therefore, it is possible to assume that when players were playing in-person, they might experienced similar changes at cognitive level but different anxious or excited timings. Moreover, according to the finding from survey results, players experienced significantly higher sense of co-presence when they were playing in-person. Therefore, we suggest sitting in the same environment could enable players to focus more on the game itself without being distracted by other factors, thus help players maintained similar cognitive experience. On the other hand, feeling more co-present could enlarge the emotional fluctuations such as excitement and anxiety for players with different identities. Therefore, SCR peaks per minute present more divergent trends over the players.

4.3. Share Physiological Data as Interaction

We conducted a field study to investigate how distributed learners react to the streaming system that presents collective HR and HRV measurements in real time. The streaming session was embedded in a lecture series lasting about 40 minutes. 48 learners in total attended the online lecture over Zoom. Eight learners volunteered to stream and share their physiological data. The system tracked BVP data from self-built wrist-worn devices with an optical sensor placed on the fingertip referring to the set-up in previous work [40, 41, 111] (Figure 3.3 (b)).

4.3.1 Real-time Analysis and Implementation

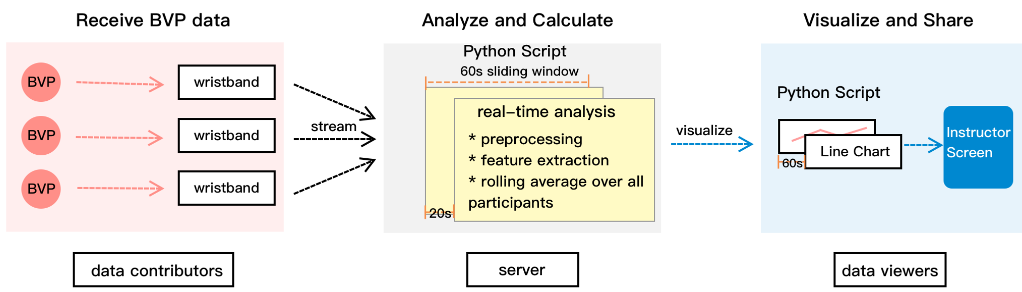

Figure 4.8 The framework of implementing real-time analysis to generate line chart online lecture. The stages are receiving BVP data, real-time analysis, and generate visualization and share.

The device sampled the BVP at 50Hz and streamed to our system server via User Datagram Protocol (UDP), which supported distributed learners to stream their data without location restrictions. Each participant's raw BVP data was passed through a 4^{th} order Butterworth low-pass filter (4 Hz). **BPM**[5], namely average heart rate, was selected as an intuitive HR indicator of excitement and anxiety. While **pNN50**[6] was adopted as an established HRV feature to reflect relaxation and sustained attention [52,93]. For each minute, the two features were averaged for each participant and were used to generate glance-able line charts right below the presenter's slide content (Figure 4.9 (right)) to avoid distracting learners from the class content. To ensure the stability of the visualization, rolling means were calculated from three data points and streamed to the visualization system. Figure 4.8 summarizes the framework of implementing real-time analysis.

5 Beats per minute

6 The percentage of adjacent NN intervals that differ from each other by more than 50 ms

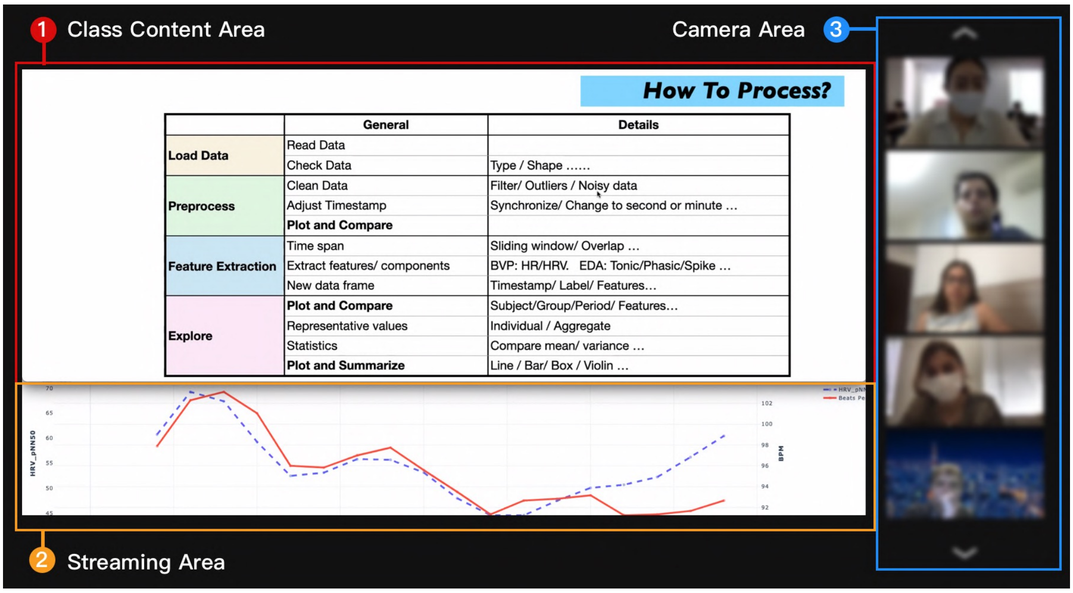

Figure 4.9 Streaming system used in the field study. Right is the interface consisted of class content area, streaming area, and camera area (Mosaic was applied to the picture for privacy concerns). The red and blue lines represented collective BPM and pNN50 respectively and both were calculated from rolling means of all data contributors every one minute.

4.3.2 Subjective Feedback

Methodology

After the streaming session, we organized interviews on 11 learners and 6 of them were data contributors while the others were viewers. During the interview, we asked participants about their perception and experience while viewing the streaming visual in the class, and to compare their experiences in class without the streaming event. We also asked participants to share their interpretations about the visuals including the meanings behind and sources of the data. We also questioned the learners who contributed streaming data about their personal feelings when wearing the device and sharing their physiological data to others. We delivered surveys after the classes with and without the streaming visualization to quantify its impact on learners. Besides demographic questions, the survey asked questions about sense of community [105] and perceived psychological engagement [109] in 7-point Likert scales.

Questions to measure sense of community were as follows (Likert scale: "1-Strongly Disagree" to "7-Strongly Agree") :

- This class helps me fulfill my needs.
- I feel like a member of this class.
- I can not get what I need in this class.
- I do not belong in this class.
- I feel connected to this class.
- I feel that I can rely on other classmates or instructors in this class.
- I feel that I matter to other classmates or instructors in this class.
- I have no friends in this class that I can rely on.
- I feel distant from other classmates or instructors.
- I have a good bond with other classmates or instructors in this class.

Questions to measure perceived psychological engagement were as follows (Likert scale: "1-Strongly Disagree" to "7-Strongly Agree") :

- I paid close attention to my classmates or instructors.
- I was easily distracted from my classmates or instructors when other things were going on.
- My classmates or instructors did not pay close attention to me.
- My classmates or instructors were easily distracted from me when other things were going on.
- I tended to ignore my classmates or instructors.

17 learners answered the survey after the class without seeing physiological data streaming, and 19 learners answered the survey after the class implemented with physiological data streaming. We notified students that joining the study was not compulsory and all of the answers would be kept confidential and not related to any type of grading.

Results and Discussions

We summarized the themes emerged from the qualitative analysis and identified feedback from data contributors and viewers. We also report survey results to triangulate the interview findings. (C:Contributor, V:Viewer)

All the participants reported they had the motivation to look at the streaming visual. Main reasons were the curiosity in others' reactions:

> *"I find it very interesting to see how people develop together."(C1))*

> *"Whether I intentionally looked the streaming was related to the instructor and the content of the course itself."(C3)*

Participants reported increasing engagement and sense of community with the streaming system. As data contributors, they could involve in the class more actively:

> *"The change of the data visualization was like the more spontaneous reaction, what's happening in class [...] when the instructor was asking us something, then I was curious if it goes up because people feel involved and then they're like, oh, I need to act now."(C1)*

and feel more connected to other distributed learners by seeing the data visualization:

> *"Putting aside how accurate the visualization were or what it meant, I felt it really cool to see that we are somehow affecting this class and visuals. [...] I felt like I was part of a class and felt like there were people like in this class that there was a presence rather than me sitting alone in the room and whatever." (C2)*

Moreover, data contributors did not report privacy concerns with the sharing mostly because of the visual was shown in an aggregated manner and perceived anonymity:

> (*"Because it feels like so many people's data and not my data alone. It's something more objective and doesn't feel violating." (C3)*).

Regarding viewers' experience, they mentioned it was a little hard to understand the system setup, such as how was the data recorded, integrated, and transformed. However, they were still interested in the trend and fluctuations of the streaming visualization:

> *"I can get a rough idea of what it is, but I don't understand what it is. But looking at the trend, I can feel the change." (V2)*

Viewers also mentioned the streaming visual enhanced their engagement and connection especially when most of the people turned off the cameras in the online class, which was a common practice of the class we studied:

> *"I could see the visuals like moving. So I can see that people are listening or people are actually there attending the class and will sort of indirectly causing me to feel like feel connected to the students." (V1)*

Survey results also showed perceived psychological engagement was slightly higher in the streaming session (mean = 3.99, sd = 0.38) than that without streaming (mean = 3.73, sd = 0.4), with a marginal significant difference (t[32]= 1.95 , (.05¡p<.10)). Although there was no significant difference of sense of community between two conditions. We will further explore in controlled lab settings to clarify the effect of streaming visualization on learners' perception of social bonds.

In accordance with our concept design, participants reported more linkage existed with the system and people who were attending the online class. Both data contributors (who were sharing their physiological data during the class) and data viewers mention the externalized heart rate data enhanced their engagement and connection especially when most of the people turned off the cameras in the online class. They also focused more on changes and fluctuations instead of absolute values and most of them prepared abstract visualizations in terms of privacy and intuitive interpretation.

4.4. Investigate Effects of Sharing Physiology

As an extension of the online lecture study, we continued to explore whether sharing physiological data could enhance entrained feelings in group activities. The practical goal of this study follows the concept of sharing physiological data as social cues. One of the aims is to explore a novel way of building connection when people chose to turn off the camera where facial expressions are no longer available.

Feedback from participants also supported the potential application of this concept. We received subjective feedback from workshop participants (those who did not attend as experiment participants could answer the questionnaire voluntarily) regarding their habits of turning on/off camera during online meetings. In total, 21 participants female:13, male:7, others:1) filled out the questionnaire and shared their reasons about why or why not turn on the camera during online meetings. Six participants never (n = 1) or rarely (n = 5) turned on the camera during online meetings. Six participants reported they sometimes turned on the camera while six participants reported they often turned on the camera during the online meetings. Only one participant always kept the camera on. Most of the participants reported their choices would change depending on the following conditions:

- Sense group dynamics: whether other group members turn on the camera; whether group members want to be engaged or connected.
- Priority of the meeting: whether the meeting is with important people; whether there is other thing to do at the same time.
- Privacy concerns: whether non-related people are in the same room; whether my private life could be exposed (e.g. joined the meeting from home).

In the study, we asked participants to turn off the camera for both sessions. We organized two online workshop sessions following one same agenda where different participants attended (see detailed information in Table 4.1). The agenda of the workshop is as follows:

- Facilitators explained the study and received the consents.
- Facilitators delivered recording devices and helped participants wear the devices before asking them sitting physically apart.
- Workshop started. One of the facilitator explained the agenda (approximately 5 minutes).
- The invited instructor delivered a short talk about how to use mixed-methods in research (approximately 8 minutes).
- Group discussion started after the instructor's talk (approximately 15 minutes).
- Workshop ended. Participants filled out the survey regarding experience (e.g.engagement and perceived entrainment [105, 109])

Table 4.1 Information about two workshop sessions.

Session	With biofeedback	Number of participants in total	Number of participants with data recorded (gender distribution)
Morning session	No	8	7 (female = 3; male = 3; others = 1)
Afternoon session	Yes	13	8 (female = 5; male = 3)

4.4.1 Real-time Analysis and Implementation

The device sampled the BVP at 100Hz and streamed to the Processing software via User Datagram Protocol (UDP). Each participant's raw BVP data was passed through a 4^{th} order Butterworth low-pass filter (4 Hz). The visualization is to mimicry heart pumps for intuitive interpretation. Therefore, filtered BVP data were directly streamed to the Processing software to control the size and color of the circle shape elements (see Figure 4.11). Figure 4.10 summarizes the framework of implementing real-time analysis.

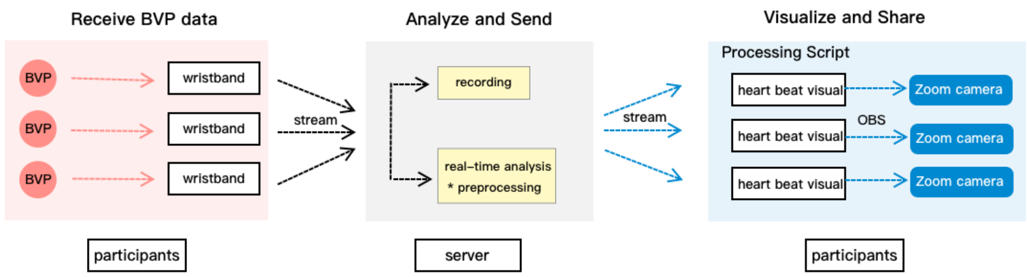

Figure 4.10 The framework of implementing real-time analysis to generate and share individual's heart beat visual during online workshop. The stages are receiving BVP data, real-time analysis, and generate visualization and share.

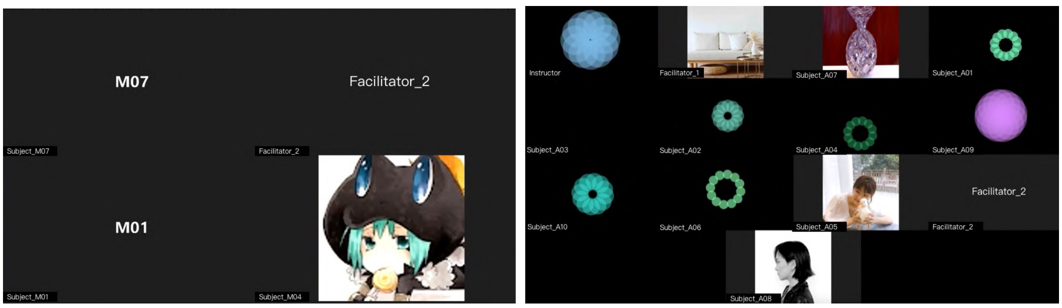

Figure 4.11 Streaming system used in the computer-mediated workshop. Left is the Zoom screen shot of morning session. Right is the Zoom screen shot of afternoon session. All the people names' have been later changed into instructor, facilitator, and subject numbers to be shown in this paper.

4.4.2 Offline Analysis

Data Collection and Dataset Description

We organized two workshop sessions and 15 participants' BVP data were recorded at 100Hz sample rate in total. Participants are students from Keio University and age between 21 40. Because we shared the zoom link to attend the online workshop publicly, we did not record the data from those who had not signed up as participants. Table 4.1 summarizes the information of participants whose data were recorded.

Offline Analysis Process

To investigate how will biofeedback influence group dynamics, we compared the trends of HRV features and quantified coordination of HRV features between morning session and afternoon session (with biofeedback provided).

We firstly adjusted timestamps and cut the dataset into the same length of the workshop. A 2^{nd} order Butterworth low pass filter (from python package, *scipy.signal*) was then used to cut high frequency noise above 3 Hz [106, 107]. HRV features were calculated every two minutes with a ten-second sliding window to capture subtle variance. For feature selection, we choose **pNN50** [7] and **RMSSD** [8] referring to the results and comparison in prior works. Both of the two features are under more influence of PSNS and could reflect high frequency variations in heart rates. Previous works have proved pNN50 could indicate relaxation level and sustained attention [47, 92, 93] while RMSSD could reflect cognitive load and stress levels [94, 95]. Following the concept of entrainment, we firstly investigated the trends of group dynamics reflected in pNN50 and RMSSD features (see Figure 4.12). We further looked into the similarity and correlation between every pair of learners during each session via the method of DTW implemented with dtw-python module[9] was used to compare quantitatively [108]. We applied an independent t-test to investigate whether a significant difference existed between the distances of every learn pairs in morning session and afternoon session. Moreover, we calculated distance between each pair's feature data by DTW every four minutes with a ten-second sliding window to observe the subtle changes in the entrainment development. The mean of the distance between the pairwise combinations was adopted to represent the entrainment of the group as a whole

7 The percentage of adjacent NN intervals that differ from each other by more than 50 ms.

8 The root mean square of successive differences between normal heartbeats.

9 `https://dynamictimewarping.github.io/python/`

(see Figure 4.14).

Results

We firstly plotted the trends of HRV features and compared between morning session and afternoon session (see Figure 4.12). Overall, average pNN50 presents an increasing trend and reached peaks around the instructor finished the short talk in both sessions. During the group discussion, average pNN50 among students in afternoon session is higher than that in morning session. Average RMSSD climbs quickly in the beginning of the workshop and fluctuates during both sessions. Overall, average RMSSD among students in afternoon session is lower than that in morning session.

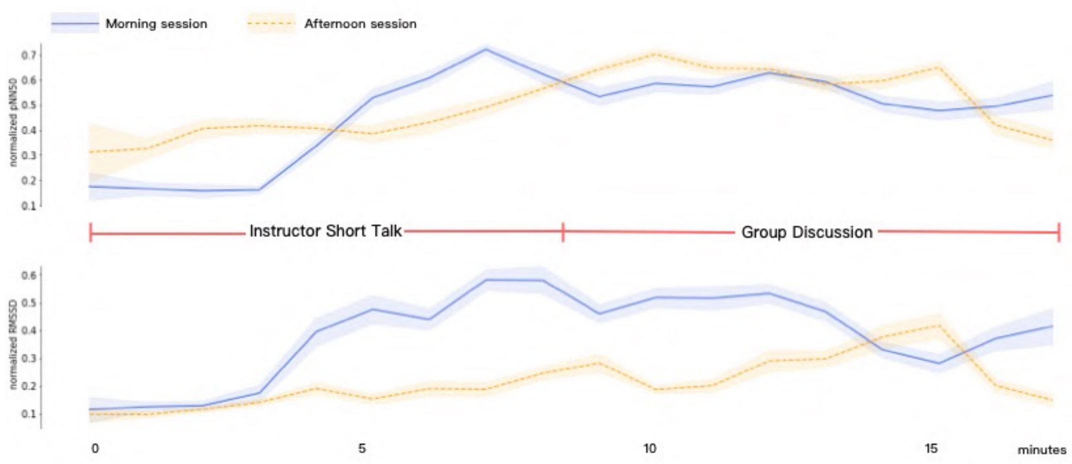

Figure 4.12 Trends of average pNN50 (top) and RMSSD (bottom) among participants during two workshop sessions: blue lines represent morning session and yellow lines represent afternoon session. The timeline of the workshop is marked in red line between two graphs.

Normalized distance calculated by DTW was adopted to quantify the similarity between every two learners' HRV features' timeseries data (pNN50 and RMSSD). pNN50 tends to be more divergent between learners in the afternoon session with biofeedback although no statistical difference is discovered (p = .079) according to t-test. However, distance between learners' pNN50 in the afternoon session declined and became smaller than that in the morning session after the group discussion started (see Figure 4.14 (top)).

Learners in the afternoon session with biofeedback (mean = 0.07,sd = 0.035) present significantly smaller distance between each other in terms of RMSSD patterns than that in the morning session (mean = 0.16, sd = 0.062), t(29.6) = 5.72, p<.001. In Figure 4.14 (bottom), a smaller distance of RMSSD in the afternoon session could also be observed.

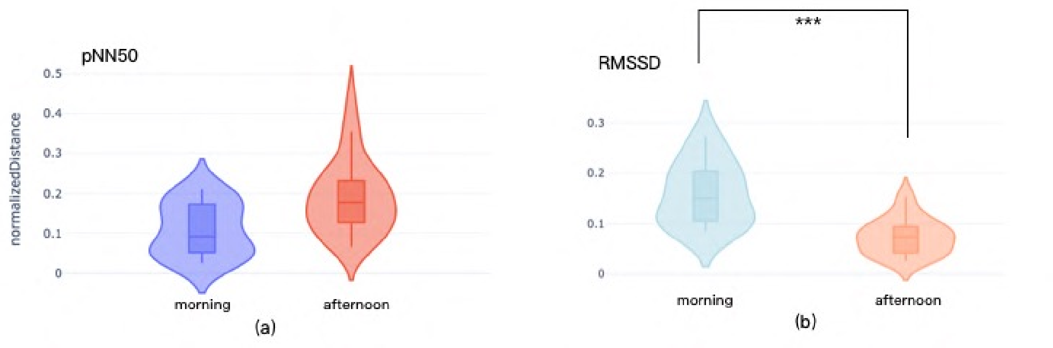

Figure 4.13 Comparison of paired distance of pNN50 (a) and RMSSD (b) between every two players during two workshop sessions. The inner box plot shows the show the minimum, first quartile, median, third quartile, and maximum values of timeseries normalized distances of pairwise physiological data's distance. The smaller the distance, the more similar the pair of timeseries data. The outer smoothed violin shape illustrates probability density. The width of the shape indicates how frequently certain values occur. ($***p$<.001).

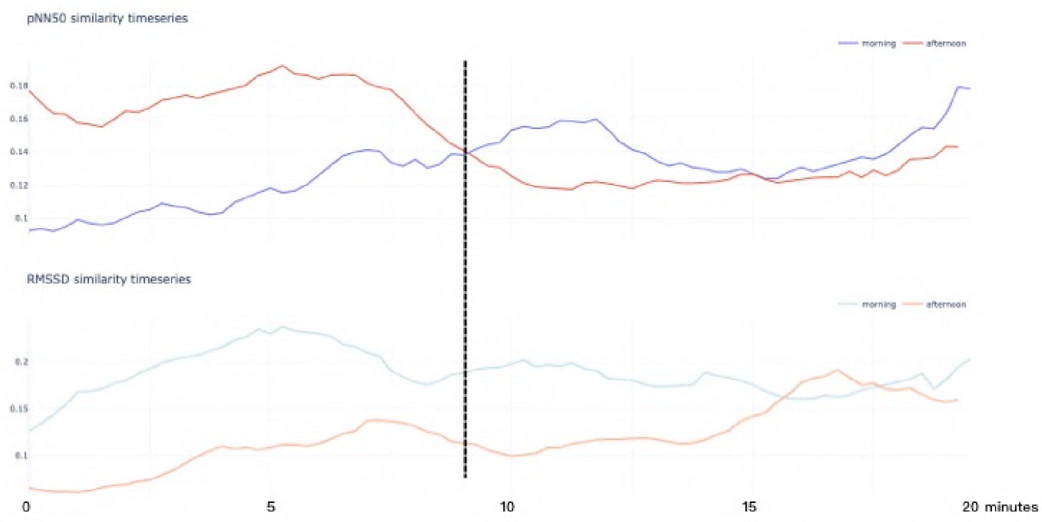

Figure 4.14 Timeseries plots of average distance among every two pairs' pNN50 (top) and RMSSD (bottom) data. The vertical black dashed line highlights the timing of when the group discussion started.

4.4.3 Subjective Feedback

Methodology

After each session, we asked only experiment participants about their psychological engagement, sense of community, and entrainment after each session by 7-Likert scale [105, 109, 110].

Questions to measure psychological engagement were as follows (Likert scale: "1-Strongly Disagree" to "7-Strongly Agree") :

- I paid close attention to my group members.
- I was easily distracted from my group discussion when other things were going on.
- My group members paid close attention to my reactions.
- My group members tended to ignore me during the discussion session.

- I feel my group members were communicating with me.

Questions to measure the sense of community were as follows (Likert scale:"1-Strongly Disagree" to "7-Strongly Agree") :

- I feel like a member of this group.
- I do not belong in this group.
- I feel connected to this group.
- I feel distant from other group members.

Questions to measure perceived entrainment were as follows (Likert scale:"1-Strongly Disagree" to "7-Strongly Agree") :

- We had similar reactions during the discussion.
- I felt excited while the other members seemed to feel calm.
- I felt sad while the other members seemed to be happy.
- The connection between me and the other members was becoming stronger.
- I felt the other group members and I were completing the task together.

Results

Average psychological engagement, sense of community, and perceived entrainment in the afternoon session are all slightly higher than those in the morning session (see Figure 4.15). However, due to the small sample size, we did not find statistical significance. Therefore, we would only use the survey results as supplementary materials.

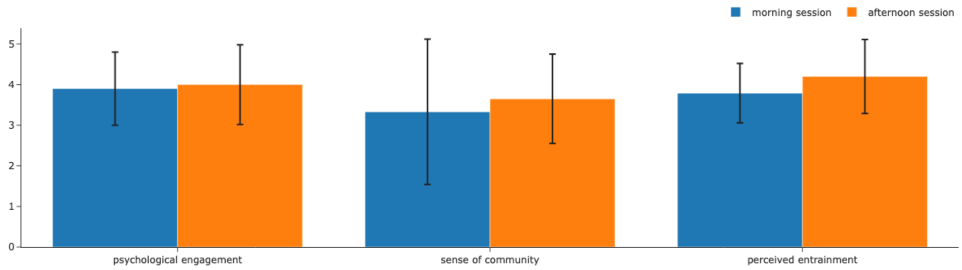

Figure 4.15 Summary of the survey results after two workshop sessions including psychological engagement (morning: mean = 3.9, sd = 0.79; afternoon: mean = 4.0, sd = 0.98), sense of community (morning: mean = 3.33, sd = 1.79; afternoon: mean = 3.65, sd = 1.10), and perceived entrainment (morning: mean = 3.79, sd = 0.73; afternoon: mean = 4.20, sd = 0.91).

Moreover, we also collected participants' feedback about the visualization implemented. Overall, the visualization could be interpreted as their own heartbeats and one participant compared the visualization with the heart rate data monitored in her other device (Fitbit). However, the ambiguity and the internet lag sometimes would make it harder to catch social hints than see face expressions directly.

4.4.4 Discussion and Interpretations

The results we presented focus on the comparison of HRV features and coordinated patterns between workshop sessions with (afternoon) and without (morning) biofeedback. As a measurement to investigate interpersonal entrainment, DTW was applied to every two participants within the group to calculate the distance between each pair's HRV feature timeseries data. Each pair's physiological data distance will be considered as a data point to reflect entrainment in group dynamics, which extends this concept from dyadic level to group level.

Physiological entrainment based on RMSSD feature shows significantly higher when we provided biofeedback that visualizes participants' heartbeats in real-time.

Participants in the afternoon session also reported higher perceived entrainment though it was difficult to find statistical difference due to the sample size. We could assume participants in the afternoon session tend to have more similar cognitive feelings since RMSSD has been proved to be an indicator for cognitive load [94, 95]. Together with the trends presented in Figure 4.12, we could notice students in the afternoon session tend to experience relatively higher cognitive load than in the morning session. One of the reasons could be learners' in the afternoon session were not familiar with the mechanism of the visualization biofeedback and need take more cognitive resources to interpret the meanings behind. However, compared with the trends of pNN50, learners in the afternoon session also tend to show more engagement in the controlled process and sustained attention [92, 93], especially during the group discussion activity. Similarly, the development of pNN50 (see Figure 4.14) in DTW indicates more correlation occurs when the group discussion started in the afternoon session with biofeedback. This finding suggests the interference of biofeedback might be more effective to keep group members more engaged when participants actively talked and interacted with each other compared with listening to the speech only.

Although, it might be hard to conclude the physiological entrainment difference was due to the introduction of biofeedback entirely. Because people's physiological states or mental states may vary during the day and the variables may not be as well controlled in the real-life workshop study. However, our findings proved the feasibility of the analysis methodology to quantify physiological entrainment to understand group dynamics and the potential to implement biofeedback as additional social cues when face-to-face communication is not available. Moreover, we found the method of plotting the time course of DTW results could directly illustrate the physiological entrainment development of group dynamics. This could enlighten another perspective of feature choices in real-time analysis to be applied in the real-time biofeedback.

4.5. Conclusion

This Chapter presented exploratory analysis (offline and real-time) on physiological data collected during in-person and online group activities.

Section 4.2 describes the exploration of HRV and EDA features to quantify group dynamics by inspecting the regressive trends from group members and calculating the pair-wise distance between each two group members through DTW. Research insights derived from the comparison between in-person and online social game experience. Results from offline analysis and subjective feedback showed more intense and diverse exciting and anxious timings existed when participants played in-person. While relatively similar RMSSD reactions could indicate more entrained cognitive activities during the in-person game.

Section 4.3 explores the idea of sharing physiological data to support online learning where social cues are usually insufficient. Subjective feedback from participants suggested sharing physiological data could increase perceived connectedness and engagement in the group activity. Section 4.4 further investigates the potential effects of this novel interaction by quantifying physiological reactions in the group with and without sharing physiological data. The results proved the concept of sharing physiological data could trigger similar physiological reactions and increase engagement yet might bring more cognitive load due to the unfamiliar information. Although the practical goal of this project is to support online group interactions, we consider research findings regarding the analysis framework could be generalized to other types of group interactions.

Firstly, we found several explainable HRV and EDA features to help researchers understand group dynamics. RMSSD, as a feature to quantify the amount of HRV, is more related to PSNS activation. Specifically, some related works in the field of neuroscience have suggested the negative correlation between RMSSD and cognitive load [94, 95]. EDA data is often used as a sensitive marker to assess emotional arousal related to SNS activation [56–59]. To extract explainable EDA features, the timings where drastic changes happened are the main focus. In this Chapter, we especially detected the SCR peaks reflecting sudden changes in skin conductance response that could be caused by increasing arousal.

We further explored the analysis methods to understand group dynamics from individuals' extracted features. Investigating trends along the development of the group event is one of the analysis methods. For RMSSD, we calculated the mean value and the variance to represent the collective experience that might be activated by PSNS. For SCR peaks, we counted the number of group members who had experienced the SCR peaks every two minutes to represent the collective arousal experience that might be activated by SNS. This exploration reminds the importance of adopting different analysis methods considering the characteristics of physiological data and features. From BVP data, continuous HRV features could be extracted by sliding windows. The dynamic changes in HRV features could reveal the physiological experience's development. Therefore, calculating the mean value among group members could help understand the collective experience. Different from BVP data, it is the timings when drastic changes happened in EDA data that reveal the event-related arousal experience. Therefore, instead of calculating the mean value among group members' EDA data, counting how many members have experienced similar drastic changes at certain timings could help quantify the collective experience. Moreover, we also explored the feasibility of DTW to compare the physiological entrainment within groups. The similarity between every two individuals in the group could be calculated by DTW as features representing pair-wise entrainment. The average value among all the pair-wise entrainment could be adopted to reflect the collective entrainment in the group.

Besides the explorations in offline analysis, we also found the potential workflow to implement real-time analysis in biofeedback systems. The workflow consists of receiving data, analyzing data in python script, and generating feedback (see Figure 4.8 as an example of the workflow). The real-time analysis could trigger feedback based on either individual data or collective data (e.g. average value among all the group members). Biofeedback through individual data could be more explicit for participants to understand the agency and connection between their data and feedback. However, it is also worth considering the privacy concerns and the possible negative feeling such as being monitored. Overall, based on the subjective feedback collected from participants, we found the real-time analysis we implemented could make the physiological data more interpretable than the raw data. Most of the participants reported the feedback could enhance the feeling of being connected. Although there are limitations due to the small sample size, the current findings suggest the potential of using physiological data to generate meaningful and reactive biofeedback to augment group interactions.

Investigations described in this Chapter are mainly based on studies with more constrained settings and relatively small group sizes. The biofeedback embedded with real-time analysis is also exploratory. The following Chapters present further explorations on large-scale groups during in-the-wild practices and advanced real-time analysis embedded in complicated biofeedback systems.

Chapter 5
Linking Group Physiology to In-the-wild Experience

5.1. Overview

The previous chapter presented explorations of analyzing physiological data to understand group dynamics in relatively constrained real-life group settings. This chapter further evaluate the analysis framework to in-the-wild live group events where researchers' facilitation was minimized. We aimed to probe feasible methods to quantify physiological reactions from large-scale groups when dataset were collected from complex and uncontrolled conditions. The dataset described in this chapter was recorded from an interdisciplinary project in collaboration with Moe Sugawa and Session House[1] – Boiling Mind Project[2].

1 https://session-house.net/

2 http://boiling-mind.org/

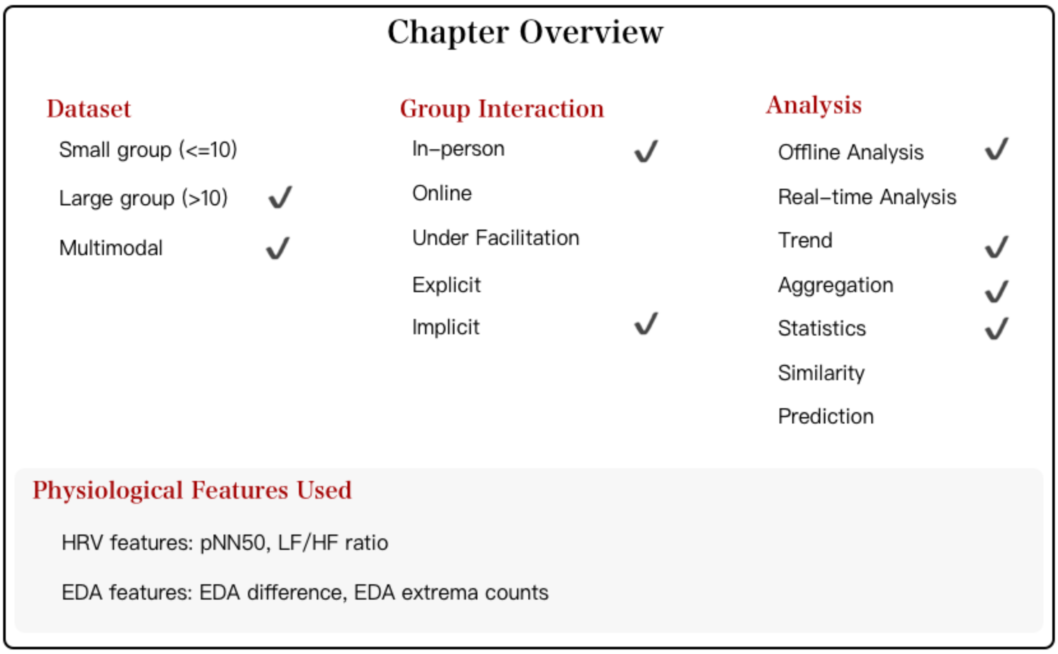

Figure 5.1 Overview of the project information in Chapter 5.

Boiling Mind is an embodied performance project combining modern dance practices, wearable sensing, and audio visual design. Both performers and researchers adopted physiological sensing as a way to explore the relationship between mind and body, invisible inner states and visible external cues [22]. For this, we followed a methodology grounded in in-situ and in the wild studies [9,13] to quantify and analyze live events. We focus on the audience in this work, introducing minimally intrusive sensing technology. We are particularly interested in the physiological changes of all audience members: are they entrained or following any rhythm? When does it happen? As an initial use-case for the work, we created a trial 15-minute dance performance [112]. Building on this, we developed a full 70-minute dance performance that was performed three times [40]. Audience physiological data (HR and EDA) were integrated into staging elements such projected visualizations and audio effects (Table 5.1 describes the key feedback implemented in the performance). This part of concept design and real-time implementation were not included in this thesis yet could be found in our previous published paper [40].

We further analyzed audience physiological data collected in the three performances to investigate and understand live audience experience. Choreography encompasses compositional design and syntactical abstractions of movement to convey an underlying meaning or idea [113]. Several works use choreographic events to predict emotional arousal measured by continuous self-report [78]. Three performances where different audience groups attended followed the same choreography, which inspires the analysis to testify whether similar reactions would exist. By comparing physiological data to the choreographic structure, our analysis revealed the link between audience collective physiology and the choreography, which contributed to a methodology to analyze and interpret in-the-wild group dynamics.

Major parts of this chapter are based on the following research papers we published:

- **Jiawen Han**, George Chernyshov, Moe Sugawa, Dingding Zheng, Danny Hynds, Taichi Furukawa, Marcelo Padovani, Karola Marky, Kouta Minamizawa, Jamie A Ward, Kai Kunze. "Linking Audience Physiology to Choreography." ACM Transactions on Computer-Human Interaction (2021).

- Sugawa, Moe, 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, pp. 1-10. 2021.

- Zhuoqi Fu, **Jiawen Han**, Dingding Zheng, Moe Sugawa, Taichi Furukawa, Chernyshov George, Hynds Danny et al. "Boiling Mind-A Dataset of Physiological Signals during an Exploratory Dance Performance." In Augmented Humans Conference 2021, pp. 301-303. 2021.

Table 5.1 Key feedback designs between the audience physiological signal and the staging elements. More details of the implementation are included in our previous paper [40]

Section	Visual Element	Sound Element	LF/HF ratio	EDA
Suits	one graph per audience member	–	value shows in the graph	value shows in the graph
Cards	one orb visual per audience member	–	control the orb's color	EDA difference controls speed of orb movement
Puppet	smoky fluid simulation	soundscape	average value controls the frequency and amount of smoke cloud's appearance	average value controls the amount of current smoke cloud and sub-frequencies into the soundscape
Romeo	one graph for Romeo	–	value shows in the graph and controls graph's color	The value shows in the graph
Growth	one wave for all members	drum sounds	average value controls the wave's color and dictates the pitch variance in the drum sounds	average EDA difference controls the height of the wave and triggers the stuttering

5.2. Choreography

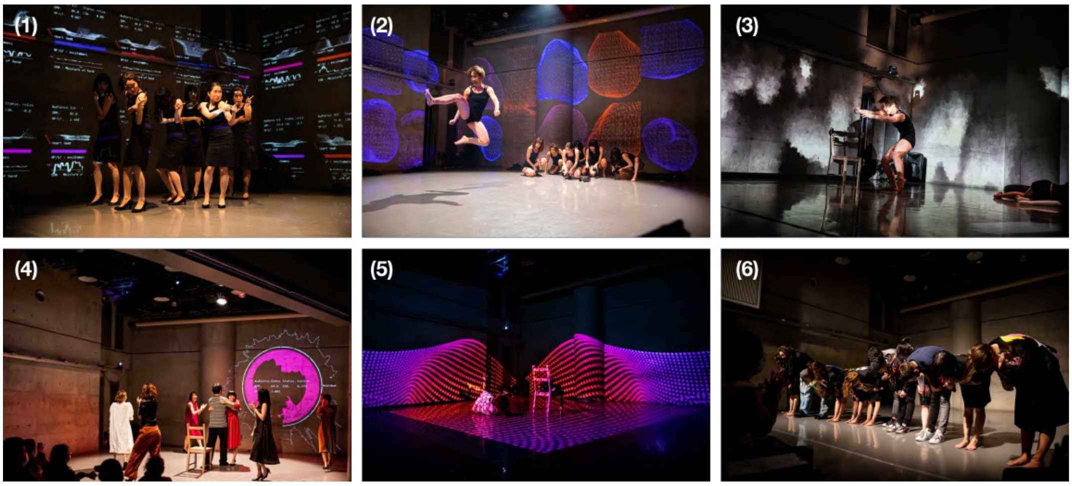

Figure 5.2 Six choreographic sections in Boiling Mind performances: (1) Section 1: Suits, (2) Section 2: Cards (Solo), (3) Section 3: Puppet, (4) Section 4: Romeo, (5) Section 5: Growth, and (6) Section 6: Curtain.

Each performance involved seven female dancers and lasted for about one hour. For analysis, we divided the recordings into six sections, each containing one or more choreographic events. These sections are shown in Fig. 5.2 and are described as follows:

Section 1: Suits. The performance starts with dancers playing the role of working women in suits and high heels trying to break out of societal pressures. At the end of this section, all dancers take off their suits and their heels. This intense movement was designed to raise the excitement level of the audience, mirroring the rhythmic and dynamic crescendo of Ravel's Boléro.

Section 2: Cards. At approximately 11 minutes into the show, the dancers engage audience members in short conversations while handing out business cards. After the dancers return to the stage, they start hitting the floor in rhythm using their heels in hand. At 17 minutes, the dynamics and gestures of Boléro reach a final peak and one of the dancers rushes to the front of the stage to perform an aggressive solo (see Figure 5.2 (2)).

Section 3: Puppet. At 18 minutes, the music turns to a more gentle and dark feel. At the same time, the previous solo dancer lays down in the center of the stage. One of the dancers brings a chair to the stage and the others gather around. All dancers start moving slowly and quietly along with the music. The chair represents everyone's position in the world. As the performance develops, each performer dances with a puppet that represents their alter ego. At the 24^{th} minute, one dancer hands the puppet to an audience member. The choreography and music work together to create a mysterious and sombre tone.

Section 4: Romeo. At 37 minutes, one of the dancers invites a man from the audience to play the role of Romeo. He is led to the stage and sat on the chair which was placed in the center of the stage. He is then asked to hug a dancer and dance together with the ensemble. The dancers start to improvise and reach out to Romeo to show they are happy to see him there. If he smiles, the dancers interact with him in an entertaining and playful way. If Romeo does not respond to the dancers accordingly, the dancers ask the rest of the audience to encourage him with applause. Some of the interactions between the Romeo and the dancers led to laughter among the audience.

Section 5: Growth. At 40 minutes, the second half of the performance develops into a deeper story. The dancers indicate the conflicting and complex feelings of instability, confusion, and joy that we all experience as we grow from childhood to adulthood. One of the dancers follows another one like a playful animal companion (e.g., a dog) willingly walking after its owner, clinging to her legs. This scene was designed to evoke the audience's sense of security and trust in being loved by others. The music for this part is quite sedated and relaxed, consisting of sparse synthesized textures and abstract rhythmic layers. As the final coda approaches, the dancers dress up as working women again but with different colorful designs embroidered into the back of their suits. This was intended to show a more positive meaning while referencing the beginning working women scene. The composed music reworked themes from Boléro into a more upbeat electronic treatment.

Section 6: Curtain. After 70 minutes there is the curtain call, when all of the dancers and crew members line up in front of the stage and bow to the audience.

5.3. Offline Analysis

5.3.1 Data Collection and Dataset Description

Self-built wrist-worn devices measuring EDA from two electrodes on the fingers, and the HR using an optical BVP sensor placed on the fingertip. The device uses an ESP32 module with Bluetooth and WiFi connectivity. It samples the BVP at 50 Hz and EDA at 4.545 Hz. In addition to the EDA and BVP, we recorded movement data using a 9-axis Bosch bmx160 absolute orientation sensor. The accelerometer and gyroscope were sampled at 50Hz, magnetometer data was not recorded. For the feedback design, only EDA and BVP data were used as input to influence visual and sound elements on the stage (see Table. 5.1

The dataset consists of audience multi-modal signals (EDA,BVP, wrist acceleration and angular velocity) over three performances. We have 98 recordings in total(male = 49; female = 49). In 1st performance, we have 34 recordings (male =17; female =17). In 2nd performance, we have 31 recordings (male =13; female =18). In 3rd performance, we have 33 recordings (male =19; female =14). By ruling out incomplete or noisy data records, we had 80 (male= 38; female=42) sets of data from the recruited participants for the HRV analysis of this project. The breakdown for each performance was: 1st, 27 (male= 12; female=15), 2nd, 27 (male=11; female=16), and the 3rd, 26 (male= 15; female=11).

5.3.2 Analysis Process

BVP

In a pre-processing step, we used acceleration data to help us identify and remove movement artifacts. To do this, we ran a peak detection algorithm on the euclidean norm of the accelerometer axes. If any peaks greater than 1.5 standard deviation were found, then we excluded the BVP data for 1s around each peak. A 2nd order Butterworth low pass filter (from python package, *scipy.signal*) was then used to cut high frequency noise above 3.5 Hz [106, 107].

HRV features were calculated every four minutes with a two-minute sliding window. For feature selection, we choose one time-domain feature and one frequency-domain feature. The HRV features were divided by mean RR intervals of each participant for normalization to remove baseline differences between individuals [114–116]. For each minute, HRV features were averaged for each participant. For each minute, pNN50 and LF/HF ratio were averaged for each participant. The data was labelled in accordance with six choreographed sections for analysis. We further mapped the timecourse of the features along the choreography to compare between the audience physiology and choreography.

EDA

Each participant's raw EDA data was passed through a 2nd order Butterworth low-pass filter from the *scipy.signal* package (0.01 Hz) [107]. We extracted EDA features by calculating the changes in EDA response which is the first derivative of the EDA data. We refer to this as **EDA difference**. Because the onset of strong emotions is typically characterised by noticeably increased sweating on the skin, we looked specifically into the timings when skin conductance drastically increased. We refer to these points as **EDA extrema**. We detected peaks and valleys of skin conductance with prominence of 1.25% of the measurement range (0 to 4095 due to 12-bit ADC) and inter-peak distance of at least 30 seconds. The EDA difference feature was normalized by MinMax scaler. Both of the HRV and EDA features was labelled in accordance with the six choreographed sections for analysis.

We counted the number of audience members who had experienced EDA extrema every minute to represent audience collective arousal feedback, which was addressed as **EDA extrema count** for describing our results and findings.

5.3.3 Results

HRV

We initially extracted the four different HRV features as detailed above: LF/HF, pNN50, SDNN, and RMSSD. The first of these, LF/HF was used in our real-time feedback system. However, we ultimately use PNN50 as a more robust measure for our main analysis. Figure 5.3 shows an example of the four features (from performance 3). Whereas pNN50, SDNN, and RMSSD all reveal a similar timecourse, LF/HF ratio presents a very different pattern (see also Figure 5.3). Therefore, we choose only one of the three similar HRV features, pNN50, for further analysis for two reasons: Firstly pNN50 is easier to interpret because it represents PSNS only, which is associated with rest, and is consequently less influenced by SNS (associated with excitement) [117–119]. Secondly, pNN50 is simple to calculate, which makes it useful as an indicator in designing a future real-time system. Figure 5.3 presents one example of the exploration process to select HRV features in the offline analysis.

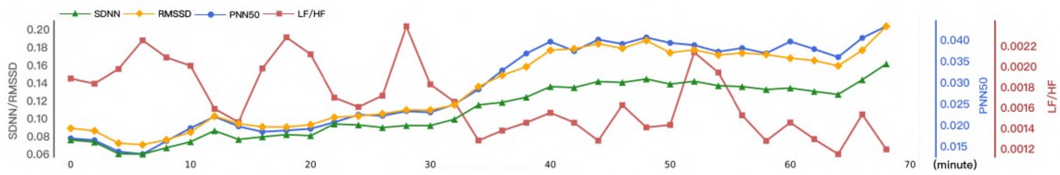

Figure 5.3 Timecourse of the four HRV features from the third performance. Significant Pearson's correlations exist between PNN50 and SDNN (r(35) = 0.99, $*p$ <.001), and between PNN50 and RMSSD (r(35) = 0.99, $*p$ <.001) - with the later correlation also reported in other works [52, 120].

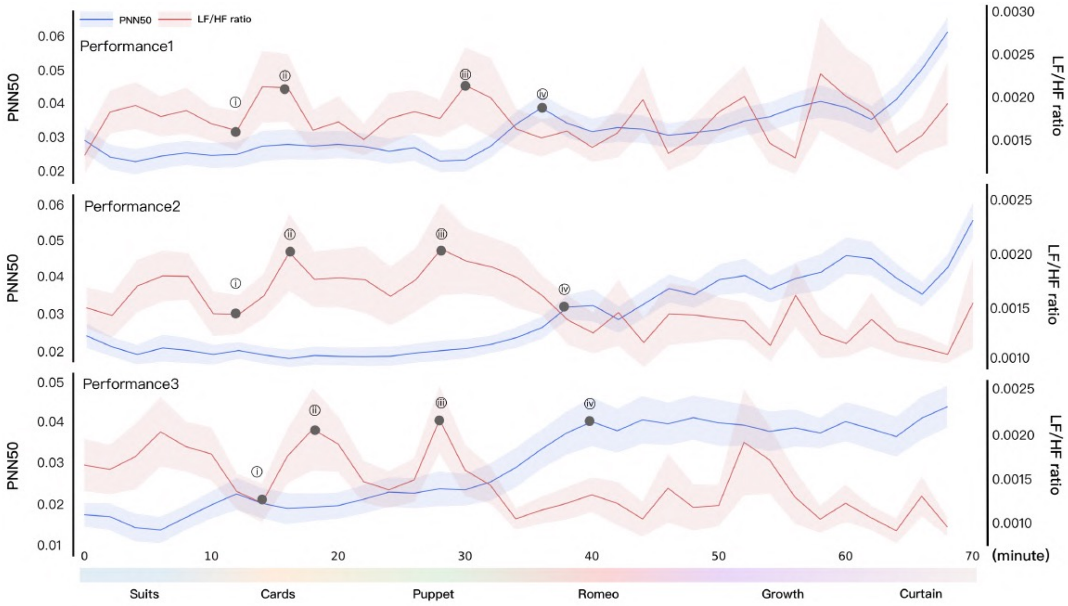

Figure 5.4 Trends of pNN50 and LF/HF ratio with noticeable turning points. The timeseries shows noticeably similar patterns of HRV feature values across the three performances. (i) highlights the decline at the end of the section “Suits”. (ii) marks peaks in the section “Cards”. (iii) highlights the peak in the section “Puppet”. Finally, (iv) marks the start of “Romeo”.

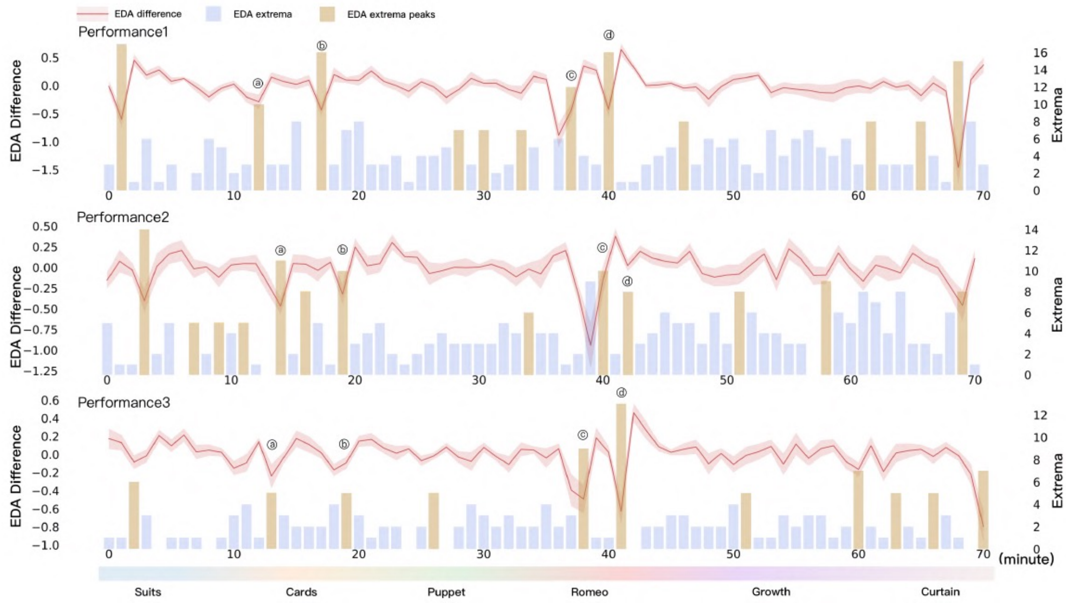

Figure 5.5 EDA difference with EDA extrema counts (bar chart). EDA extrema peaks (highlighted in yellow) are selected as being over 1.5 standard deviation from the total EDA extrema counts, compared to EDA extrema within two or more minutes. The timeseries shows noticeably similar patterns of EDA feature values (EDA difference declined with outstanding EDA extrema peaks) at certain scenes across the three performances. The scenes are marked as a (the start of the section “Cards”), Around b (the end of “Cards”), c (the start of “Romeo”), and d (the end of ”Romeo”).

We inspected the timeseries of the average LF/HF ratio and PNN50 over all audience members for each performance (see Figure 5.4). The timeseries shows noticeably similar patterns of HRV feature values across the three performances. For example, LF/HF ratios decline at the end of the section ”Suits” and start rising at the start of the section ”Cards”Then it first drops, and peaks again at around 30 minutes. The PNN50 is low at the start, but rises steadily throughout the performance. However, there is a sharper and drastic increase between the end of the section ”Puppet” and the start of ”Romeo”.

By mapping the trend and variance of three audience group’s physiological data along the choreography, we found the collective physiological data could reflect the overall unfolding of the choreography and highlight the notable moments during the performance (see Figure 5.4). We inspected the timeseries of the average LF/HF ratio and pNN50 over all audience members for each performance (see Figure 5.4). The timeseries shows noticeably similar patterns of HRV feature values across the three performances. For example, LF/HF ratios decline at the end of the section "Suits" and start rising at the start of the section "Cards". Then it first drops and peaks again at around 30 minutes. The pNN50 is low at the start but rises steadily throughout the performance. However, there is a sharper and drastic increase between the end of the section "Puppet" and the start of "Romeo". An increase of the pNN50, being closely linked to PSNS, is associated with relaxation [47, 121]. This trend could indicate that the audience was easing into the performance as it progressed.

HRV Scene Aggregate

Table 5.2 Descriptive statistics of HRV and EDA features over six sections in Boiling Mind performance.

		LF/HF	*PNN50*	*EDA difference*	*EDA extrema counts*
		Mean(SD)	*Mean(SD)*	*Mean(SD)*	*Mean(SD)*
Performance 1	Suits	.0017 (.0009)	.025 (.018)	.023 (.096)	4.18 (4.71)
	Cards	.0019 (.0014)	.026 (.021)	-.049 (.109)	5.75 (5.23)
	Puppet	.0018 (.0013)	.027 (.017)	.026 (.051)	4.29 (2.31)
	Romeo	.0015 (.0009)	.035 (.020)	-.231 (.203)	8.20 (5.59)
	Growth	.0017 (.0011)	.037 (.017)	.007 (.071)	4.54 (2.06)
	Curtain	.0019 (.0024)	.061 (.023)	-.344 (.462)	6.75 (6.24)
Performance 2	Suits	.0016 (.0007)	.022 (.014)	-.020 (.106)	4.00 (3.82)
	Cards	.0016 (.0008))	.020 (.011)	-.058 (.098)	4.13 (3.87)
	Puppet	.0018 (.0012)	.021 (.011)	.016 (.065)	3.82 (2.01)
	Romeo	.0015 (.0007)	.031 (.017)	-.200 (.234)	5.20 (4.09)
	Growth	.0014 (.0007)	.038 (.015)	.037 (.084)	4.85 (2.24)
	Curtain	.0014 (.0011)	.049 (.020)	-.193 (.300)	4.25 (3.30)
Performance 3	Suits	.0021 (.0012)	.015 (.013)	.060 (.100)	1.64 (1.75)
	Cards	.0019 (.0012)	.019 (.015)	-.012 (.090)	2.88 (1.36)
	Puppet	.00183 (.0011)	.022 (.017)	.016 (.039)	2.47 (1.46)
	Romeo	.0015 (.0007)	.035 (.026)	-.130 (.207)	3.20 (3.35)
	Growth	.0015 (.0007)	.038 (.019)	-.002 (.047)	2.96 (2.58)
	Curtain	.0012 (.0005)	.044 (.025)	-.246 (.306)	2.75 (3.10)

As a further analysis, we aggregated the timeseries to produce statistics for each of the six main sections. A repeated measures ANOVA with a Greenhouse-Gessier correction was used to investigate the correlation and variance. For the post-hoc

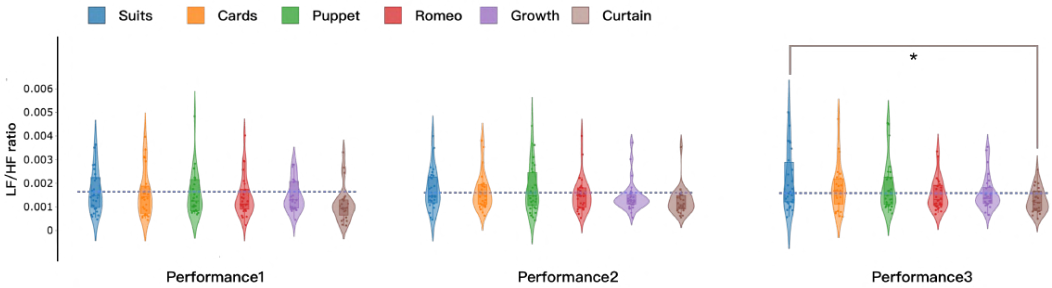

Figure 5.6 Distribution of LF/HF ratio in six sections of three performances. The violin plots illustrate probability density, while individual observations are the dots within the violin graphs. The horizontal blue line represents the average LF/HF for each performance. The only significant pairwise difference is between Suits and Curtain in performance 3 (*p<.05).

.

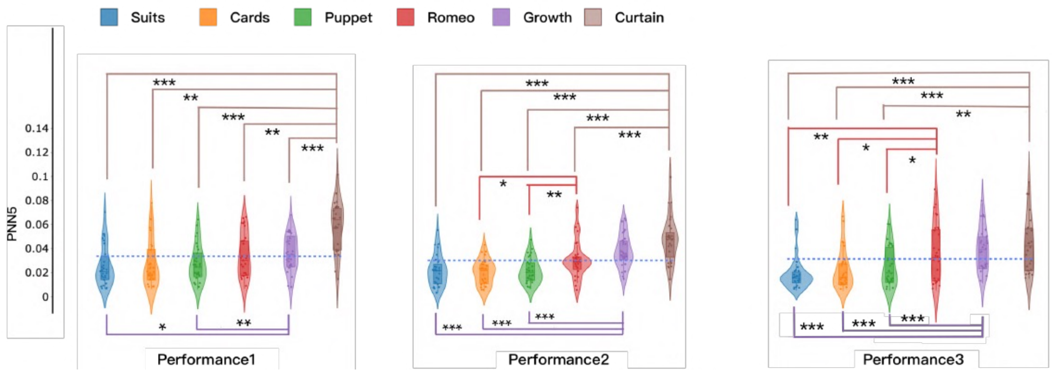

Figure 5.7 Distribution of PNN50 in six sections of three performances. The violin plots illustrate probability density, while individual observations are the dots within the violin graphs. The horizontal blue line represents the average PNN50 for each performance.The vertical lines drawn between two graphs indicate the significant levels of pairwise comparison results (*p<.05, **p<.005, ***p<.001).The pairwise comparison shows distinguishable separation between the first half (before Romeo) and the second half (from the end of Romeo).

.

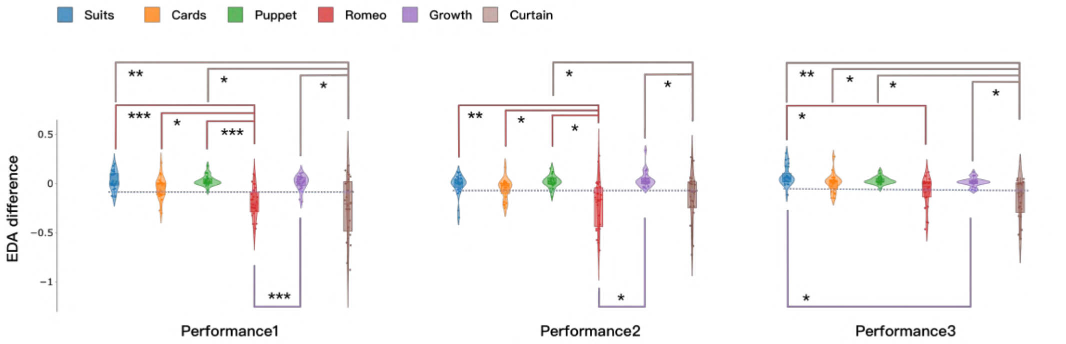

Figure 5.8 Distribution of EDA difference in six sections of three performances. The violin plots illustrate probability density, while individual observations are the dots within the violin graphs. The horizontal blue line represents the average EDA difference for each performance.The vertical lines drawn between two graphs indicate the significant levels of pairwise comparison results ($*p<.05, **p<.005, ***p<.001$). Romeo and Curtain show significant differences to the other sections.

tests, we applied Bonferroni correction to prevent the inflation of type-I errors.

There were no statistically significant differences for LF/HF across the six sections over performance 1 (F(1.73, 44.93) = 0.351, p = .675) and performance 2 (F(3.15, 81.85) = 1.16, p =.33). Significant differences were present in performance 3 (F(3.37, 84.34) = 4.84, p = .003). In the post-hoc analysis we found that significant differences only existed between "Suits" (M = .0021, SD = .0012) and "Curtain" (M = .0012, SD = .0005) with p = .027 each. Since these two sections marked the beginning and end of only one performance, this effect is likely an anomaly. These results also depicted by Fig. 5.6.

When analysing the mean PNN50 value, we found statistically significant differences between the six sections in performance 1 (F(2.81, 73.04) = 30.39, p ¡ .001), performance 2 (F(3.10,80.70) = 34.26, p ¡ .001), and performance 3 (F(2.56,64.08) = 18.40, p ¡ .001). (Full descriptive statistics are provided in Appendix **??**, Fig. 5.7 depicts the distributions and pairwise comparisons.)

EDA

We inspect the EDA response using our two features: average EDA difference and EDA extrema counts (Fig. 5.5). The timeseries reveals large changes at the beginning of each performance when the lights go off and the music starts, as well as at the end when the performers take a bow. Throughout the performances there are also common changes at around 13 minutes, marked in Fig. 5.5 as (a), 19 minutes (b), 37 minutes (c), and 41 minutes (d). (Note that EDA extrema is shown in bar chart form to highlight that, unlike the other features, it represents a discrete count rather than an average.)

EDA Scene Aggregate

We calculated an aggregate pairwise comparison of EDA difference distributions between the 6 sections for each of the 3 performances (shown in Fig. 5.8). After Bonferroni correction, we found that both Romeo and Curtain are statistically different to the other sections. According to a repeated measures of ANOVA with a Greenhouse-Gessier correction, mean EDA difference values differed in a statistically significant way between the sections in performance 1 ($F(1.61, 41.93) = 13.41$, $p < .001$), performance 2 ($F(2.17,49.97) = 8.15$, $p = .001$), and performance 3 ($F(2.16,49.56) = 11.34$, $p < .001$). (Full descriptive statistics are provided in Table5.2.)

5.4. Subjective Feedback – Audience

5.4.1 Methodology

We used online questionnaires after each performance to gather audience feedback. These were accessible using a QR code on flyers handed out to each attendee. Responses were encouraged but not mandatory.

The questionnaire assessed demographics, cultural background (how often do you visit theater/dance performances), and performance specifics (enjoyment of performance). Free-text answers were given to the specific question of “how much did you feel like participating in the performance”, as well as general opinion on the piece.

The full list of questions in the questionnaire delivered to the audience were as follows:

- How much did you enjoy this performance overall? (Likert scale: "1-Not at all" to "9-Very much")
- How much did you enjoy the visualization/ music/lighting/dance? (Likert scale: "1-Not at all" to "9-Very much")
- Compared to other performances, how much did you feel participating in the performance? (Likert scale: "1-Nothing" to "9-Strongly") and Why did you have this feeling? (free-text answer)
- Which staging elements excited you most? Why did you have this feeling? (single choice: visualization, music, lighting, and dance)
- Please leave your opinions freely on this performance. (free-text answer)

We received questionnaires from 35 participants in total (self-identified as male= 16; female=18; prefer not to say=1). Since the questionnaires were not completed by all participants, we consider these answers as supplementary information.

5.4.2 Results: Audience Qualitative Feedback

Among the 35 respondents, 30 reported to experience of watching dance (every week: N=3; every month: N=8; every year: N=19). For the open-ended questions, we categorized the feedback and reported as follows:

Participation in the Performance

A lot of feedback described strong feelings of participation compared to previous experiences. The simple knowledge that the audience was sensed might have played a part in this:

> *"I was not sure if my heart rate really affected the visuals, but I was excited when thinking my heart rate was being measured. I felt like I was on stage at that time."*

"The display of physiological data and the link between color and excitement were impressive and I felt I participated in it."

Some participants reported a feeling of connection between audience and dancers:

"Lighting and visuals changed in response to the audience's sensors, and I enjoyed the two-way interactions in this performance".

However, others found the system confusing:

"The music and visuals changed as our excitement changed. However, the lighting was a little difficult to notice."

And some even felt a disconnect between how they felt and what they saw:

"Sometimes the visuals from the sensor data matched my excitement while sometimes they did not match."

Memorable Moments

Generally, participants considered the visualization, music, and dance as intriguing and meditative:

"Whenever I watch their dance, something refreshing my memory happens. This time I had this feeling as well."

Several audience members shared memorable moments:

"I felt that music, sound, rhythm, and breaks tended to be the switches of excitement. During Bolero's gradation and explosion, rhythm and dance were connected closely."

"Dancers eye contacts when they using high heels to hit the floor were cool."

"I may be more excited in quiet and dark moments than when I'm feeling something intense will happen. I thought dancing the chair and the scene of the Japanese song were wonderful."

Sense of Unity

When asked about their free opinions, most of participants mentioned they experienced a strong sense of unity between audience and dancer during the performance:

> *"The abstract visual expression was very beautiful in connection with the dance. I felt that my senses were integrated with the dance through this indirect media."*

And among audience members:

> *"I can feel not only my own sense of participation, but also other audience's reaction reflected. I was able to realize the sense of unity between the audience, which is usually hard to feel."*

However, one audience member doubted the need to enhance the sense of unity between dancers and audience suggesting that audience reactions may vary a lot due to different compositions of audience and this could make quality control harder:

> *"In dance performances, "today's audience's feeling" and "sense of unity" seem to be less important to me. If the music and lighting change depending on the audience of the day, the impression of the work will change accordingly."*

5.5. Subjective Feedback – Dance Team

To understand dancers' personal experience during the performance and attitudes towards the collaboration process, we conducted a focus group with five dancers which draws on the approach used by Huskey et al. [122].

5.5.1 Methodology

The semi-structured focus group was conducted via video conferencing three months after the performance when we had identified initial research insights.

One of the dancers was the facilitator. We encouraged the dancers to talk freely using the following pre-prepared questions as a guide:

1. Share your experience and feeling when you saw the visualization, lighting, and the change of music triggered by the audience reactions. (Did you notice anything interesting, shocking, or disappointing?)
2. Was this experience different from previous performances?
3. What do you think about this collaboration? Share some experience of your memory about the collaboration.

All participants discussed in Japanese and videos were recorded for later transcription. We translated the transcripts and categorized the qualitative feedback.

5.5.2 Results: Dance Team Feedback

In general, dancers followed the choreography as they rehearsed without influence from the audience-derived visualizations. Two reasons were given for this. First, they had to focus on the performance itself with little time to care about changes in visualizations:

> *"In the scene of Bolero, I need imagine the train's passing by during my dance and was not able to pay attention to visualizations until the scene changed",* (D1).

> *"When I was waiting for my turn, audience heart rate displayed was very exciting. After I started dancing, though it was fun to see that, I could not spare my attention to the changes. I think we need more times to get used to it",* (D3).

The second is that they were sometimes confused by the meaning of the visualizations:

> *"I felt it was tough to balance between something researchers want to show and something dancers want to show. It was difficult for me to fully understand the visuals' meanings and the formal performance day came",* (D1).

> *"I did not understand the meaning when the lighting started to flicker. I did not see it as audience heart rates and it did not change that obviously",.* (The audience BPM data was mapped to the intensity of the lighting changes.) (D2)

Three dancers said that there were moments when they could sense audience reactions and one dancer were even influenced to adjust their movements:

> *"I was aware of the visualizations when it came to the Romeo scene while the data used in generating the feedback loop was from the chosen audience",.* (D1)
>
> *"It was very easy to see when audience felt more excited, but it was less noticeable for the calmer scene and I felt that audience's feelings did not change from the visualizations", (D2).*
>
> *"When I danced with a puppet, I noticed the coloring of the visualizations were blue and I tried to dance intenser and faster, even hit the floor more painfully to get audience more sympathized and aroused",* (D4).

The dancers also thought there could be more space for improvisation where they could dance according to audience reactions but it would be more difficult in terms of the choreography (D1, D2, D4). To solve this, D2 mentioned they could *"predetermine some triggers and reactions"* accordingly during certain moments instead of improvising throughout the whole performance.

5.6. Interpretation and Discussion

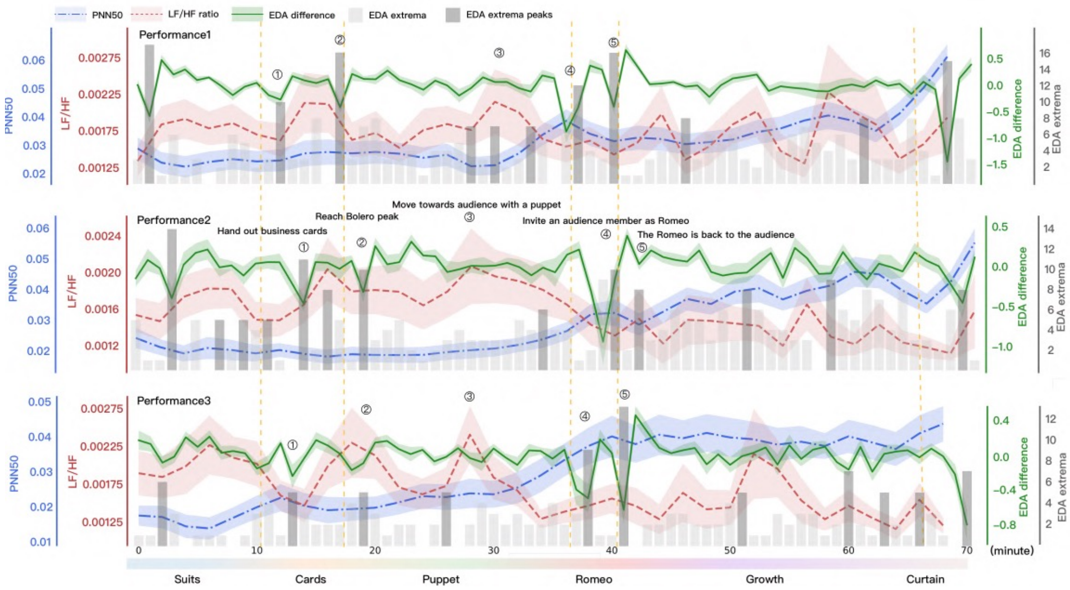

Figure 5.9 The change of HRV features (Left Y scale) and EDA features (Right Y scale). EDA extrema peaks (highlighted in dark grey) are selected as being over 1.5 standard deviation from the total EDA extrema counts,compared to EDA extrema within two or more minutes.Five key moments are marked as 1 (hand out business cards), 2 (reach bolero peak), 3 (move towards audience with a puppet), 4 (invite audience member as Romeo),and 5 (the Romeo is back to the audience).

5.6.1 Connection Physiological Data to the Choreography

The Performance Timeline

According to the choreographer, the first half of each performance (before the section "Romeo") was designed to directly engage the audience and to elicit emotional reactions that might be more easily captured by the system. This included direct interactive elements (e.g., "Cards") as well as tense musical rhythms (e.g., "Puppet"). In the section "Romeo", a direct and close interaction between one selected audience member was designed to elicit extreme involvement and empa-

Table 5.3 Physiological changes at five notable moments marked in the Figure 5.9.

Moment	Choreography elements	**Changes in HRV**		**Changes in EDA**	
		LF/HF ratio	*PNN50*	*EDA Difference*	*EDA extrema*
①	Direct interaction, dancers to the audience, short time (5 s)	–	–	Abrupt drop	Outstanding peak
②	Music builds up, aggressive solo, strong rhythm by high heels	Noticeable spike	–	Abrupt drop	Outstanding peak
③	Dancers moving towards the audience, long time (20 s)	Noticeable spike	–	–	–
④	Direct interaction, one audience member to the stage	–	High Level	Abrupt drop	Outstanding peak
⑤	The audience member back	–	High Level	–	Outstanding peak

thy. The second half (from the start the section “Growth”) was designed to be less interactive and more reflective. In contrast to the previous half, “Growth” did not involve any interaction between dancers and the audience to help the audience digest the piece and to reflect on their own experiences. Hence, it provided space to focus on the inherent message of the performance. Several members of the audience reported that they experienced this as a process of ”*meditation*”.

Accordingly, the pNN50 shows a general rising trend across all performances (see also Figure 5.9) and a significant increase comparing with first half sections (see Figure 5.7). An increasing of the pNN50, being closely linked to PSNS, is associated with relaxation [47, 121]. This trend could indicate that the audience were easing into the performance as it progressed. The aggregated pNN50 value also presents statistically significant differences between the six sections in performance 1 ($F(2.81, 73.04) = 30.39$, $p < .001$), performance 2 ($F(3.10,80.70) = 34.26$, $p < .001$), and performance 3 ($F(2.56,64.08) = 18.40$, $p < .001$). (Full descriptive statistics are provided in Table 5.2, Figure 5.7 depicts the distributions and pairwise comparisons.)

Notable Moments from the Data

From Figure 5.9 it is hard to identify a clear long-duration trend of LF/HF over the three performances. The lack of any significant inter-scene difference supports this. However spikes of LF/HF ratios appear when audiences were subjected to relatively long scenes with an intensive crescendo (e.g. as the Bolero dance peaks ②, or when the Puppet is moved towards the audience ③). As an indicator of the balance between SNS and PSNS activity [97], LF/HF ratio could implicitly imply stress [46], anxiety [88], or excitement [98, 99, 123]. Even though the interpretation of the LF/HF ratio is controversial [96], it is still possible to explain the changes with cautious consideration of the recording contexts [97]. Since the trend of LF/HF ratio does not synchronize with that of pNN50 (an established measurement of PSNS), we are inclined to believe that these changes of LF/HF ratio were mostly due to increased arousal under the influence of SNS activity. This is supported by previous works, on the physiological responses to music, that showed significant LF/HF increases during exciting, fast-tempo music [98, 99].

Across all performances there is a clear steepening in pNN50 starting from the end of Puppet and Romeo. This is also reflected, in part, in the pairwise scene aggregate results of Figure 5.7, where this change is significant for performances 2 and 3. Since an increasing pNN50 can also reflect engagement in the controlled process, and sustained attention [92, 93], we connect this rising of pNN50 to the choreography design to trigger a sense of security and reflection.

When interactions between audience and dancers were direct and intense, both abrupt drops of EDA difference and outstanding peaks of EDA extrema counts can be observed in Figure 5.9 (①,②,④). As an index of emotional activation, EDA could reflect arousal regardless of valence types [56, 57, 124, 125]. Previous audience studies have connected EDA to engagement [5] and shock-effects during the performance [7]. In our study, we would assume observed EDA changes reflect audience surprise ①, and excitement ②, and tension ④ modulated by the choreography. Among those scenes, EDA difference presents most drastic declines around the Romeo scene ④. As a collective measurement for the audience group, EDA extrema counts are also highest around the Romeo scene. This suggests that inviting someone on stage (to play Romeo) may trigger sudden tension among the remaining audience (i.e. in sympathy, or in anticipation that they might be invited next).

5.6.2 Interaction with the Audience's Physiological Feedback

In the BoilingMind performance, the interaction design between the audience and the spatial places is inspired by somaesthetics [26, 126]. The way audience members interact with the dancers and staging elements was limited and implicit instead of actively controlling or replying to the system. On one hand, both dancers and the audience mentioned the feelings of being sensed and connected subtly led them to a sense of unity and connection. On the other hand, the ambiguous and introspective atmosphere may be the reason for the unclear interpretation of the feedback. The dancers reported focusing more on their movements than on the changing audience feedback. One reason could be the lack of the perceived agency within the interaction because the effects were triggered by the audience's physiology instead of the dancers' own responses as in previous works [29, 126, 127]. Another reason could be due to their lack of experience with the novel technology. Despite this, there were some scenes where the dancers responded strongly to the audience feedback, such as when they moved faster in order to elicit a change in the coloring.

Drawing on this, it would be useful to explore a tighter integration between this technology and dance by incorporating the practice of audience feedback earlier in the rehearsal process. This would familiarise the dancers with the system and allow them to explore more nuanced and interesting responses to unexpected feedback. The choreographer and the artistic director suggested some focal points for future improvement. One focus is to investigate different forms of aesthetic interaction that might generate a clearer feedback loop between the audience's physiological reaction and the improvisations. Another focus is to improve the audiences' feeling of comfort during the performance - enhancing confidence and trust in the performance environment, both artistically and with the technology. Creating a suitable environment benefits the audience experience, their enjoyment of the performance, as well as enhances the potential to obtain better quality data for research.

5.6.3 Lessons Learned

The work presented here is primarily practice-led, where research methods, contexts and outputs involve a significant focus on creative practice [27, 128, 129]. Based on our investigation of the audience's physiological data and the co-design process, we summarize the lessons learned for both HCI researchers and performance artists.

For HCI practitioners interested in performance and audience interaction, our approach explores an effective way to collect, analyze, and interpret audience experience during live performance. Live dance performance is a useful in-the-wild scenario to explore interaction paradigms that move away from the individual and towards interactions in larger-scale groups. Although academic research is usually conducted as goal-oriented while artistic practice is more process-driven [129], both teams converged around the common goal to enhance the invisible link between dancers and audience through performance.This co-design process led to a series of novel performances and large-scale physiological data collections from the audience. Our exploration of the dataset reveals a link between the choreography and the audience's physiology. PNN50, being closely related to PSNS activation, shows a general rising trend and a significant increase from the second half sections. We found PNN50 could be a reliable and robust indicator of the audience's tension and relaxation during the performance. Moreover, LF/HF ratio, EDA difference, and EDA extrema could reflect the audience physiological reaction (e.g. excitement, surprise, anxiety) elicited by choreographic elements such as strong rhythms and direct audience interaction. Our findings suggest the potential for a more holistic view on understanding and quantifying audience experience by cross-mapping choreography and physiology.

For performance artists, our research opens a viable method to incorporate audience physiological data within a live performance. Based on our post-analysis of the audience physiological data and the feedback from the dancers, we provide suggestions about choosing HRV and EDA features for live feedback. EDA difference is well-suited to gauge audience reactions in real time since it is sensitive to short and sudden changes in arousal like shock effects. HRV features may be used to reflect the audience's moment-to-moment experience or long term growth of emotional arousal. PNN50 in particular is a robust measure for visualizing a sustained change in engagement, or a shift from tension to relaxation.

The choreographer and the artistic director suggested some focal points for future improvement. One focus is to investigate different forms of aesthetic interaction that might generate a clearer feedback loop between the audience's physiological reaction and the improvisations. Another focus is to improve the audiences' feeling of comfort during the performance - enhancing confidence and trust in the performance environment, both artistically and with the technology. Creating a suitable environment benefits the audience experience, their enjoyment of the performance, as well as enhances the potential to obtain better quality data for research.

5.7. Conclusion

This chapter investigates how to link physiological data collected from large-scale group (98 audience members) during in-the-wild conditions to group experience. The exploration of the dataset and collected qualitative feedback enabled us to discover how does the audience physiology became entrained under the effect of choreographic design (The dataset together with the sample code for analysis is available for researchers under this link: https://osf.io/sypz4/). Therefore, mapping collective physiology to the group event time course (e.g. the overall story theme and interactive elements) could help researchers and practitioners with different background to gain a multi-modal and holistic view of live group experience.

Based on our offline analysis results, we found out characteristics of certain HRV (pNN50 and LF/HF ratio) and EDA features (EDA difference and EDA extrema) which could imply for feature choices in real-time analysis. We further discussed the uses of selected features for understanding group dynamics and designing real-time biofeedback in lessons learned. Through the reproducible approach described in this chapter, we are progressing towards understanding and reflecting group dynamics. The following chapter presents our further explorations of investigating and influencing the invisible connection between group members.

Chapter 6

Influence Physiological Experience in Group Dynamics

6.1. Overview

The previous chapter presented our approaches of analyzing large-scale physiological data to understand group dynamics during in-the-wild experience. This chapter extends the exploration by investigating the methodologies to detect and share specific physiological experience during real-life group events. The dataset used for developing real-time detection algorithm was collected in lab environment. Dataset for offline analysis to investigate group dynamics was collected from in-the-wild concert where real-time detection algorithm was embedded in the biofeedback system. Both of the analysis were parts of an interdisciplinary project in collaboration with Yan He – Frisson Waves Project[1].

1 https://cybernetic-being.org/works/frisson-waves/

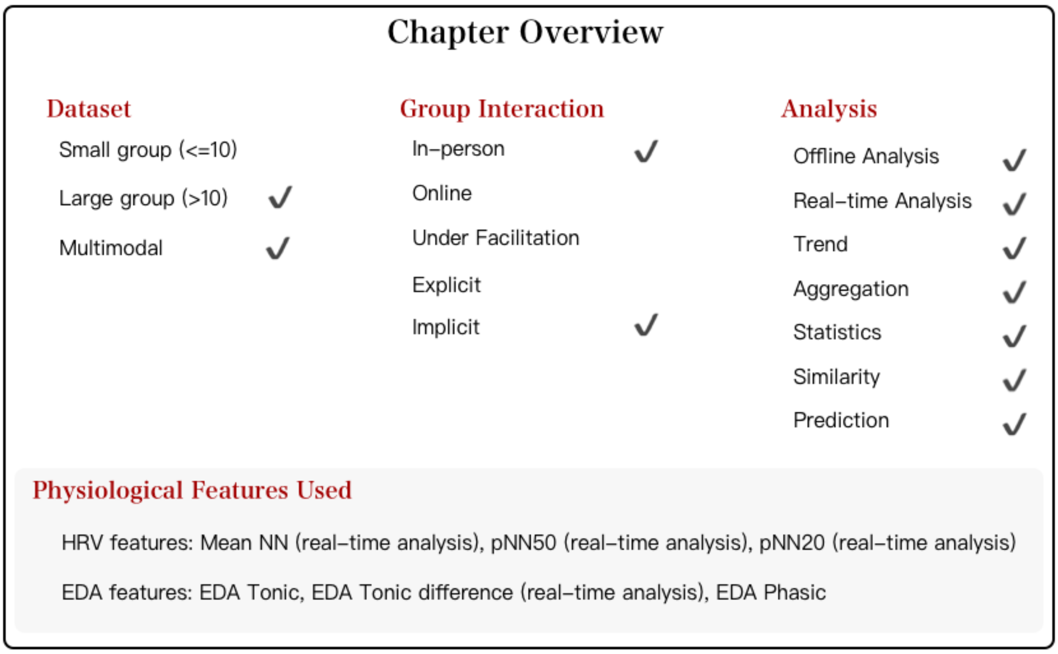

Figure 6.1 Overview of the project information in Chapter 6

Frisson Waves is a project to detect, trigger, and share particular sensational feelings (frisson) in a wave-like pattern over audience during music performances. Frisson is a psycho-physiological phenomenon commonly described as having goose-bumps, or feeling shivers down one's spine, that can be triggered from external stimuli such as music or intense emotions [130,131]. Frisson usually happens when music deeply resonates with people and those who experience frisson tend to feel higher degrees of pleasure than those who do not experience frisson [132]. However, there is an obvious individual difference regarding the frisson experience. Some people may frequently experience frisson while some people may never have this feeling. Therefore, we assume sharing and triggering frisson sensations with biofeedback may increase the aesthetic experience, especially for those who never experience frisson naturally. Figure 6.2 shows the biofeedback system we developed and implemented in real-life concerts.

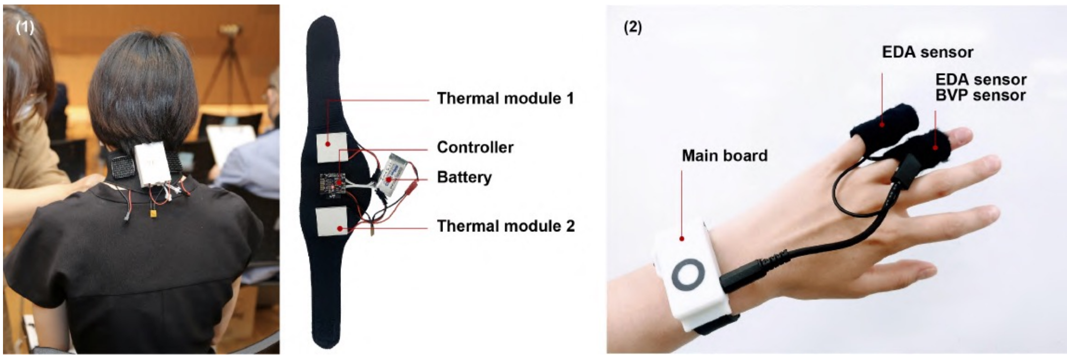

Figure 6.2 Thermo-haptic neckband and wristband used in the biofeedback system implemented in the piano concerts.

The system was developed iteratively over the course of four studies: an in the wild study piloting the design requirements for frisson feedback during a live performance, a lab study validating a method for automatic frisson detection, a lab study evaluating the frisson inducing device, and a final in-the-wild study evaluating the complete system during a live performance. This Chapter firstly describes the process and results of developing real-time frisson detection algorithm for biofeedback actuation. Most related works explain frisson as a type of physiologically contagious emotional arousal which could be indicated through EDA and HR/HRV related features under the influence of ANS [133–137], which provided neuroscience basis for building a model detecting frisson events through physiological data especially EDA related features. Section 6.2 describes the lab study where we collected BVP and EDA data with frisson events labelled and the process of building and evaluating the frisson detection model.

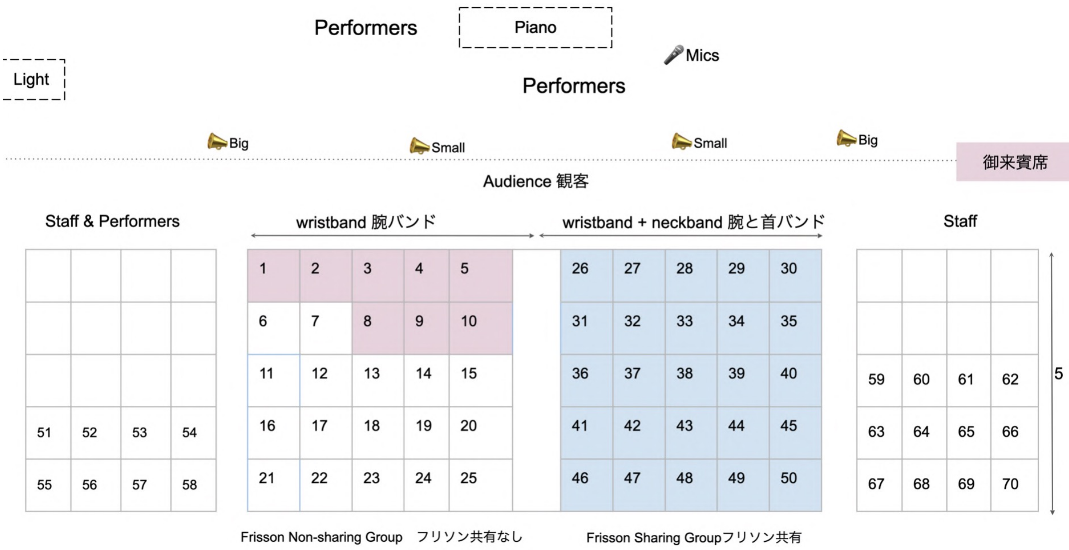

Figure 6.3 Illustration about the seats' plan for Sharing and Non-sharing group.

At the final concerts, the audience were divided into frisson Sharing group and Non-Sharing group (see Figure 6.3). By analyzing the collected physiological data, we could further investigate whether different or common physiological reactions exist between audience groups with and without frisson triggering. According to previous work, higher physiological entrainment among audience members in classical music concerts could be positively related to their aesthetic experience and social connectedness [138, 139]. We also further quantified the physiological entrainment and compared between two groups. The offline analysis and results were described in Section 6.4.

Major parts of this chapter are based on the following research papers we published:

- Yan He, George Chernyshov, **Jiawen Han**, Dingding Zheng, Ragnar Thomsen, Danny Hynds, Muyu Liu, Yuehui Yang, Yulan Ju, Yun Suen Pai, Kouta Minamizawa, Kai Kunze, Jamie A. Ward, "Frisson Waves: Exploring Automatic Detection, Triggering and Sharing of Aesthetic Chills in Music Performances." Proceedings of the ACM on Interactive, Mobile, Wearable and

Ubiquitous Technologies 6, no. 3 (2022): 1-23.

- Yan He, George Chernyshov, Dingding Zheng, **Jiawen Han**, Ragnar Thomsen, Danny Hynds, Yuehui Yang, Yun Suen Pai, Kai Kunze, and Kouta Minamizawa, "Frisson Waves: Sharing Frisson to Create Collective Empathetic Experiences for Music Performances." In SIGGRAPH Asia 2021 Emerging Technologies, pp. 1-2. 2021.

6.2. Real-time Frisson Detection

6.2.1 Data Collection and Dataset Description

To collect and label the dataset for model training, we conducted a 30-minute lab study and recruited 33 (self-identified as female = 17; male = 16) participants to collect labeled EDA and BVP data. There were three sessions and the order was counterbalanced.

- Session 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 provoking chills [140].
- Session 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".
- Session C is a five-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.

During the study, participants were first explained the definition of frisson and were asked to press the button for reporting frisson experience either elicited by music or thermal stimuli. Figure 6.5 shows one subject's example EDA data patterns in frisson events (natural and triggered) that are similar enough to be used equally in the later model training. Once the study began, participants filled out the demographic questionnaire followed by the explanation of frisson while

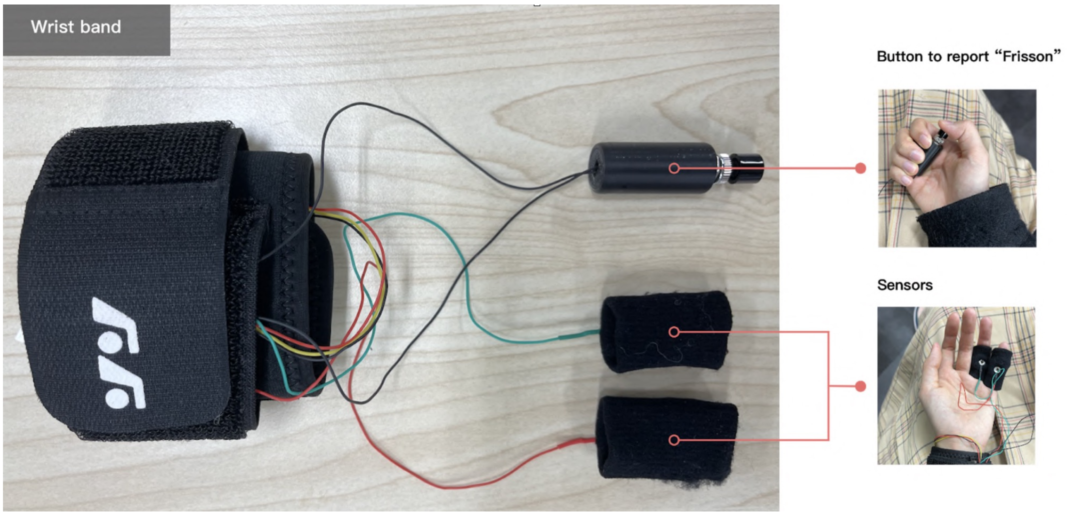

Figure 6.4 Device used in the lab study to collect participants' EDA and BVP data. The device is customized with a self-report button to label frisson events for later frisson detection model training.

the investigators helped them put on the wristband with a frisson-report button (see Figure 6.4) 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.

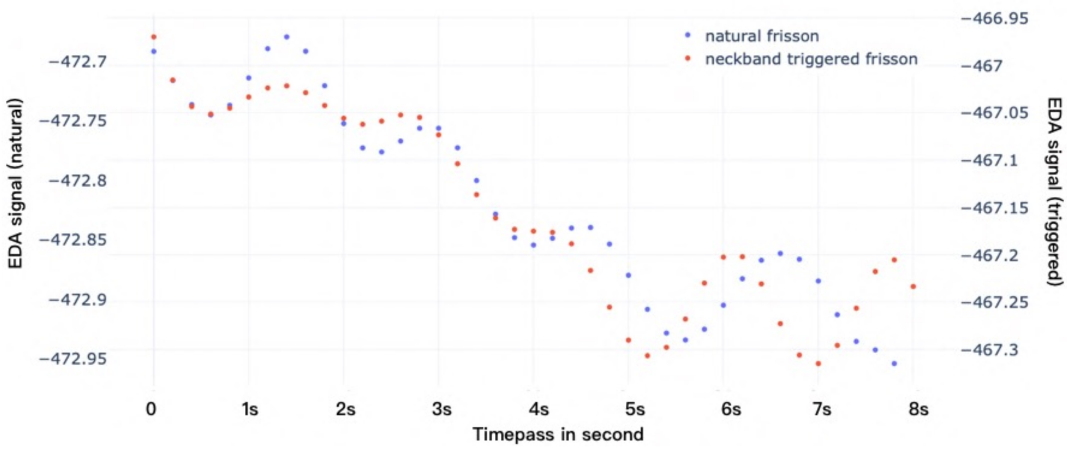

Figure 6.5 Comparison between natural and triggered frisson events.Similar trends of filtered EDA data for natural and triggered frisson events lasting for around eight seconds could be visually inspected.

6.2.2 Model Training and Evaluation

We removed the participant's data if the data was either too noisy to process or few frisson events were reported, which left us 19 participants' data in total. Each participant's raw EDA data was passed through a 2$^{\text{nd}}$ order Butterworth low-pass filter (0.5 Hz). And each participant's raw BVP data was passed through a 4$^{\text{th}}$ order Butterworth low-pass filter (4 Hz). We extracted four EDA features and three HRV features using Neurokit2 [100] from the filtered signals. We later used scikit- learn library to compute feature importance and train frisson detection model[2] [141]. According to the feature importance by a Random Forest Classifier, we selected the following features and normalized the values to remove individ-

2 https://scikit-learn.org/stable/

ual differences: **EDA-Tonic[3], EDA-Phasic[4], Tonic-diff-60[5], Tonic-diff-30[6], MeanNN[7], pNN50[8], pNN20[9].**

We used 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 [88, 142]. Our model was trained using the features extracted from a sliding window of 1 minute moving every 1 second. (Note that during the real-time detection stage, to ensure sufficient data in the event of data loss, 2 minutes of data is sent, from which only 1 minute is ultimately classified at a rate of every 10 seconds.) The window was labeled as a frisson event if the button was pressed within the window.

Physiological data of both the natural frisson and triggered frisson events were analyzed and plotted. Firstly, we visually inspected the trend and change of filtered EDA data and found the two types of events were similar enough to be both considered as frisson experiences. Figure. 6.5 shows one subject's example of frisson events (natural and triggered) and the trend of filtered EDA signal recorded. Moreover, we trained a trial model using the data from three subjects combined who reported frisson in both natural sessions and triggering sessions. The model was trained on the data in the natural session and tested on the triggering session with an accuracy of 80.42% as an initial result. Both of the results supported our concept. We further developed our model using all the data from 19 participants 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. 6.6 illustrates the overall performance of the model.

3 EDA Tonic component value

4 EDA Phasic component value

5 Change of EDA Tonic component value in 60 seconds

6 Change of EDA Tonic component value in 30 seconds

7 Average of normal sinus beats' interbeat intervals (NN)

8 Percentage of adjacent NN intervals that differ from each other by more than 50 ms

9 Percentage of adjacent NN intervals that differ from each other by more than 20 ms

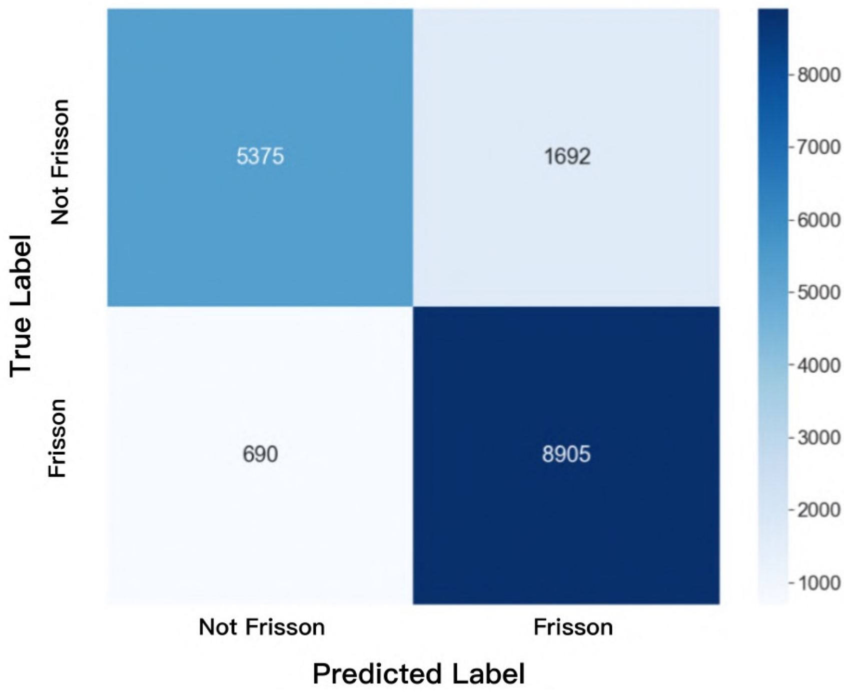

Figure 6.6 Confusion matrix for the frisson detection model with LOPO-CV. A row represents an instance of the actual classes whereas a column represents an instance of the predicted classes. Each number represents the sample numbers falling into each quadrant. The overall sensitivity (true positive rate) is 92.81% and the overall specificity (true negative rate) is 76.06%.

6.2.3 Real-time Model Implementation

Figure 6.7 summarizes the framework of the real-time frisson detection model's implementation. The server software controls each device via a TCP/IP network connection. The server records all the data from the wristbands and manages the data processing and recording. The server starts a python script for each device that extracts the EDA and HRV features necessary for the frisson classification model and runs the classifier every 10 seconds for each device. If an occurrence of frisson is detected, the python script reports it back to the server. Then the server commands all the neckbands adjacent to the participant who had just experienced frisson to activate and apply cold feedback to their wearers. The implementation of real-time frisson detection model proved the feasibility of integrating real-time analysis to biofeedback systems for augmenting entrained experience.

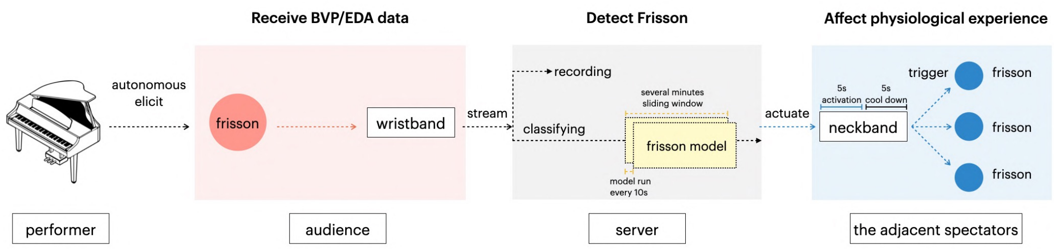

Figure 6.7 The framework of implementing real-time frisson detection model in the biofeedback system. The stages are receiving BVP/EDA data, real-time frisson detection, and affect physiological experience.

6.3. Concert Information

Five professional musicians, each with roughly 20 years of experience, curated and composed a musical program based on their artistic interpretations to evoke the audience members' aesthetic responses. In the following, we describe the concert program.

The concert consisted of three sessions summarized as follows:

- Session 1 was a 20-minute interactive ambient music performance titled "Reflections on Chopin Prelude Op.28 No.15" which was based around two electronic artists generating sounds with laptops and synthesizers from the audience's real-time heartbeats and frisson physiological signals using custom-built patchers in Max/MSP. These patchers function by receiving data from the server via OSC (open sound control) protocol, which is then distributed by the artists to several filters, triggers, and sound modifiers. One violinist and one pianist performed with the electronic artists together with the audiences' physiological feedback loop, which transformed the performance from a more structured classical work into a semi-improvisational performance

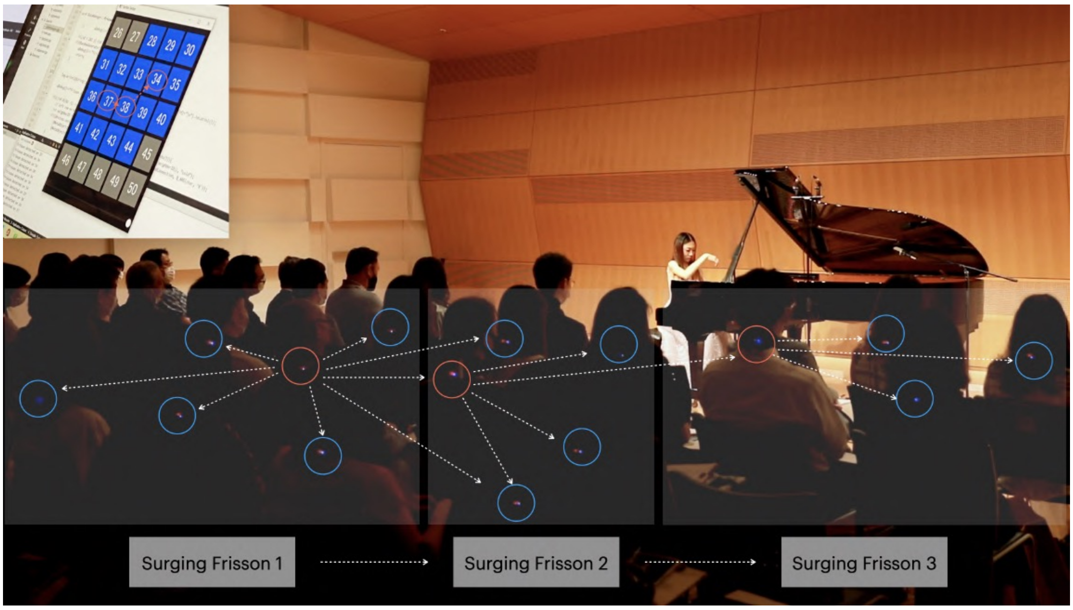

Figure 6.8 Example of frisson transmission effect matched with seat number in the Frisson Wave Concert. The real-time frisson detection model is running on the computer server.

piece. We composed this piece to express the lost feeling in the gloomy periods and bring the audience implicitly together with us to complete this piece. We programmed this piece as a process from a weak repetitive beat to a clear and brighter sentence with a growing and arising end.

- Session 2 was a 21-minute classical piano program of "Beethoven: Sonata No.30 Op.109" [10]. In the opinion of the pianist, this sonata, composed by Beethoven in 1820, is very compact both spiritually and technically. It has a wide variety of emotions, including melancholy, joy, and a feeling of grace. It sometimes portrays resentment and conflict as well. In particular, the third movement, which begins with a naive and romantic theme and consists of six variations and a coda, expresses various human emotions skillfully such as deep love and conflict hidden in Beethoven's innermost feelings.

- Session 3 was an 18 minutes classical piano program of two pieces by Frederic Chopin. The first was "Nocturne Op.27 No.1" [11]. This nocturne, composed in 1835, was written in a deeply sorrowful tonality in C sharp minor. A theme consisting of a wide-range chord in the left hand and a simple melody that echoes above it in the right hand. Chopin's delicacy has changed from the mysterious grace to the dramatic appearance of Mazurka in the middle part and the sudden appearance of Mazurka comes as if to express the national feelings towards Chopin's home country Poland, with the piece then returning to the main theme quietly. It revealed his love of his homeland and the resentment and anxiety that dwelled somewhere in his heart. The second piece was "Preludes Op.28 No.18 24" [12] [13]. These preludes, which were completed on Mallorca in 1838, were influenced by Bach's equal temperament and were composed with one song each, covering all 24 tonalities. The length and difficulty of each piece are different. Although there is no unity between the songs, the characters of each tonality are expressed in

10 https://www.youtube.com/watch?v=8JZGiY--2LM

11 https://www.youtube.com/watch?v=wuL7UC2glJM&t=1813s

12 No.22:https://www.youtube.com/watch?v=ejUG_nAEQKM

13 No.24: https://www.youtube.com/watch?v=QHcEH2Rliko

a delicate, graceful, and bold way. That is typical of Chopin's works, and the harmony of the songs before and after is well maintained. Each piece is a straightforward projection of Chopin's music and feelings for life. It is a collection of preludes that skillfully expresses human emotions. Chopin composed these pieces to cover the travel expenses for his escape to Mallorca with George Sand.

6.4. Offline Analysis

6.4.1 Data Collection and Dataset Description

48 audience members in total attended the concert (female=28; male=19, 1 other or preferred not to say) between 19 and 83 years (mean =38.53, sd=15.09). The audience registered voluntarily through a concert poster posted on social media. We included the concert's time, location, and music programs in the delivered poster. When they came to the concert, we prepared flyers for each participant to illustrate the concert and together with the consent forms about data usage and photography before they entered the hall. COVID-19 infection prevention measures were implemented in this concert such as regulating the distance between all participants as well as sanitizing all the wearable devices before and after usage. The participants could stop wearing the device or leave the hall anytime they want.

48 audience members were able to choose to sit in one of two groups after understanding the different experiences during the concert: Sharing and Non-sharing groups. The Sharing group consisted of 24 audience members wearing our neckbands and wristbands (see Figure. 6.8). The Non-sharing group of 24 audience members wore only the wristbands.

6.4.2 Analysis Process

To investigate the impact of sharing frisson experience on connectedness, I conducted offline analysis by quantifying physiological entrainment within each group. Physiological entrainment occurs when the "physiological activity between two or more people" becomes associated or interdependent" [4], which could be a feasible metric to quantify the experience of physiological connectedness. After removing noisy and incomplete datasets, we have the EDA data from 9 audience (Non-sharing: 4, Sharing: 5) in the first session, 16 audience (Non-sharing: 7, sharing: 9) in the second session, and 16 audience (Non-sharing: 7, Sharing: 9) in the third session. Each participant's raw EDA data was passed through a 2^{nd} order Butterworth low-pass filter (0.5 Hz).

We conducted the decomposition on filtered EDA data into EDA Tonic and EDA Phasic and extracted SCR peaks via one python package – Neurokit2 [14] [100]. We normalized the two components to remove individual differences via the MinMax scaler. EDA Phasic, as one event-related EDA feature [57, 60, 61], were averaged for each participant and mapped along the music development in each session. Additionally, we counted the number of audience members who had experienced SCR peaks every minute to represent audience collective arousal feedback.

Moreover, following the prior work of Gashi et al. [67] we processed DTW to calculate the distance between pairwise timeseries data (normalized EDA Tonic and EDA Phasic) in each group [67]. Smaller distance could indicate more physiological entrainment that usually occurs when the "physiological activity between two or more people" becomes associated or interdependent" [4]. Therefore, we consider this measurement could be used as a feasible metric to quantify the "bond" in group dynamics. Since DTW is a pairwise analysis method to quantify entrainment, measures were calculated for every pair within each group and compared between groups.

6.4.3 Results

We inspected the timeseries of the EDA Phasic for three concert sessions with different themes (see detailed concert information in Section 6.3). Through comparison among the three sessions, we found SCR peaks counts in Session 1 were less than those in Session 2 and Session 3 generally and were also more evenly distributed than those in Session 2 and Session 3. In Session 3, we could notice an obvious increase of SCR peaks counts around a, b, and c.

14 `https://neuropsychology.github.io/NeuroKit/`

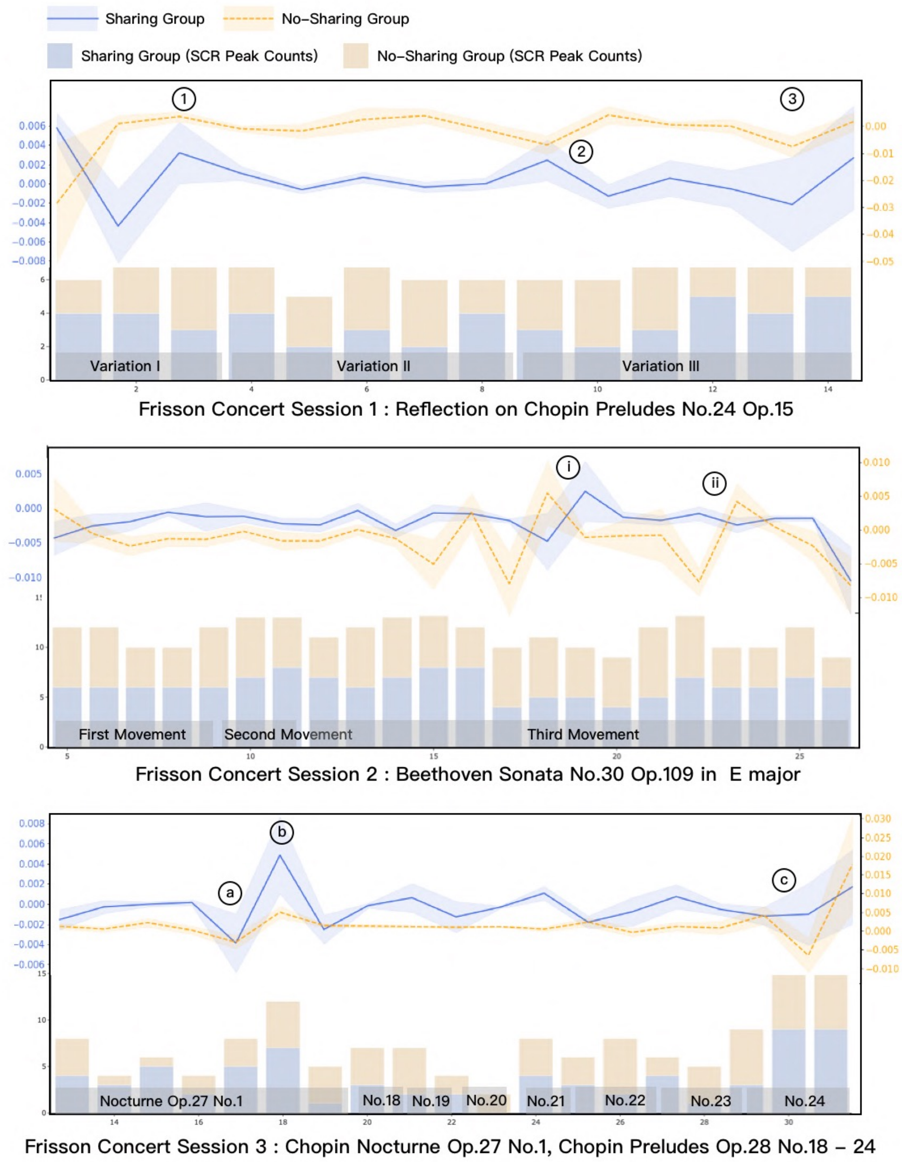

Figure 6.9 Trends of EDA Phasic and SCR peaks counts over three concert sessions. ①, ②, and ③ mark the rise around the transitions between two variations or the end of variation in the first session. ⓘ and ⓘⓘ highlight the frequent fluctuations when the music has unexpected change and building up dynamics in the second session. ⓐ and ⓑ mark the peaks and rapid changes from the middle of Nocturne Op.27 No.1. ⓒ highlights the sudden surge of EDA Phasic and SCR peaks counts in the No.24 of Chopin Preludes Op.28.

The average EDA Phasic in Session 1 tends to show a more stable trend than those in Session 2 and Session 3. In Session 2, average EDA Phasic fluctuated frequently in the third movement of Beethoven Sonata No.30 Op.109 in E major, especially around i and ii. In Session 3, the average EDA Phasic, especially the average EDA Phasic of Sharing Group, fluctuated intensely around a and b.

Moreover, by comparing the average EDA Phasic trends of Sharing and Non-sharing groups, we found they had similar reactions in general. For example, EDA Phasic kept stable at certain periods (e.g. Variation II in Session 1, Second Movement in Session 2, No.18 No.23 in Session3.) while fluctuated frequently sometimes (e.g. i in Session 2 and b in Session 3). However, we noticed Sharing Group has more vigorous dynamics.

Figure 6.10. presents the distribution of accumulated distances normalized with the signal length between each pair within the group. According to the t-test results, we found the normalized accumulated distances of EDA tonic in the Sharing group (mean = .19, sd = .11) was significantly larger than that in the Non-sharing group (mean = .10, sd = .06),$t(56) = 3.51$, $p<.001$. However, the normalized accumulated distances of EDA phasic in the Sharing group (the first session: mean = .003, sd = .002; the second session: mean = .003, sd = .002) was significantly smaller than that in the Non-sharing group (the first session: mean = .008, sd = .002; the second session: mean = .005, sd = .002) for both the first session, $t(15) = -4.50$, $p <.001$, and the second session, $t(56) = -2.81$, $p <.05$.

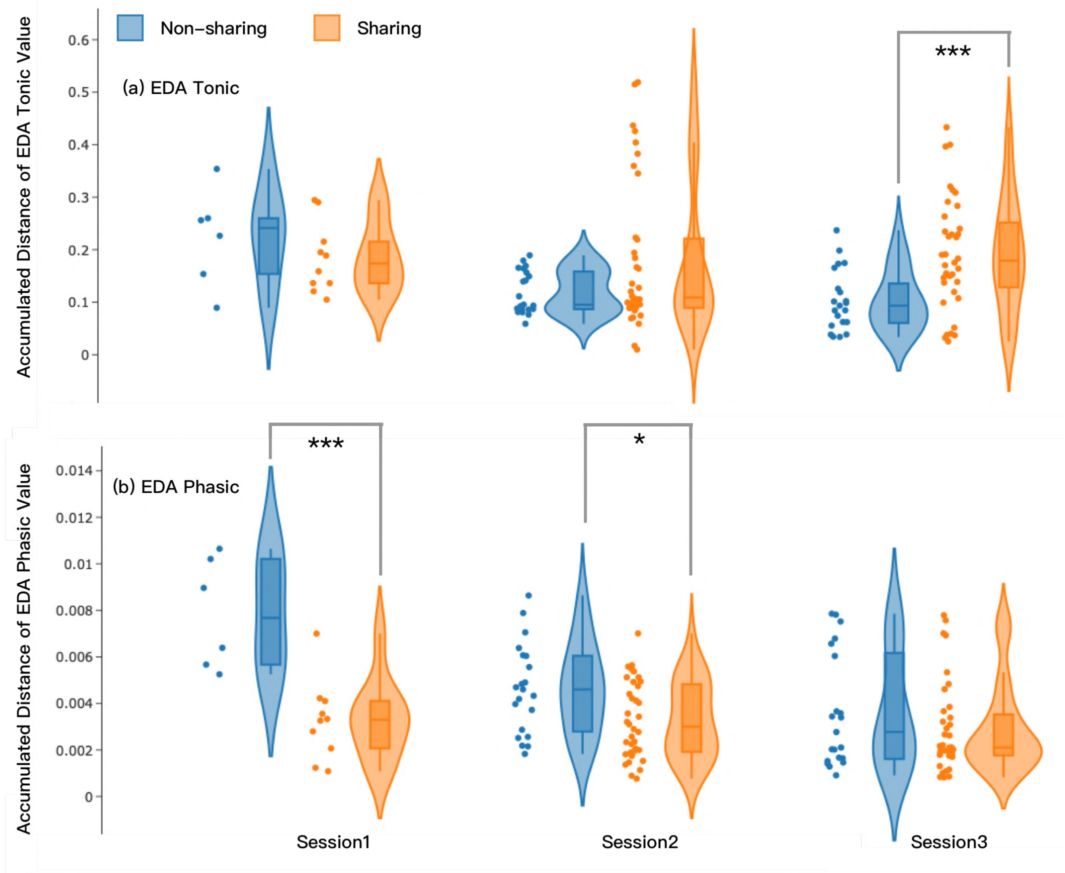

Figure 6.10 Comparison of paired similarity of EDA tonic (a) and EDA phasic (b) between every two audience members in two groups. The inner box plot shows the show the minimum, first quartile, median, third quartile, and maximum values of timeseries normalized distances of pairwise physiological data's distance. The smaller the distance, the more similar the pair of timeseries data. The outer smoothed violin shape illustrates probability density. The width of the shape indicates how frequently certain values occur. In the third session, the distances of EDA tonic in the Sharing group was significantly larger than that in the Non-sharing group. While, in the first two sessions, the distances of EDA phasic in the Sharing group was significantly smaller than that in the Non-sharing group. ($*p$<.05, $**p$<.005, $***p$<.001)

6.5. Subjective Feedback – Audience

6.5.1 Methodology

We invited audience to answer the questionnaires handed out at the reception after each performance to gather audience' subjective feedback. Responses were encouraged but not mandatory. The questionnaire assessed demographics, frisson occurrence, and concert experience through likert scale.

Questions related to frisson occurrence in the questionnaire delivered to the audience were as follows:

- Before the concert, were you familiar with frisson/aesthetic chills. (Yes, No, Maybe)
- In general, how often do you experience frisson? (Likert scale: "1-Never" to "5-Very Often")
- How often do you experience frisson in relation to music? (Likert scale: "1-Never" to "5-Very Often")
- How often did you feel frisson during the concert? (Likert scale: "1-Never" to "5-Very Often")
- I feel frisson (aesthetic chills) often when listening to music. (Likert scale: "1-Disagree Strongly" to "7-Agree Strongly")
- I have felt very intense frissons during this concert. (Likert scale: "1-Disagree Strongly" to "7-Agree Strongly")

Questions related to concert experience and connectedness in the questionnaire delivered to the audience were as follows (Likert scale: "1-Never" to "5-Very Often") [138]:

- I enjoyed the music performances.
- I found the concert more engaging than other similar performances I attended.
- I felt connected with people sitting around me.

For audience members who were wearing the neckband (Sharing group), we asked questions related to the experience wearing the neckband as follows (Likert scale: "1-Never" to "5-Very Often"):

- I experienced more frisson than in a normal concert performance from the neckband.
- I found the neckband decreased my enjoyment of the performance.

Moreover, to eliminate the effect of interpersonal relationship, we also asked audience members to report the acquaintances sitting around by checking on Figure

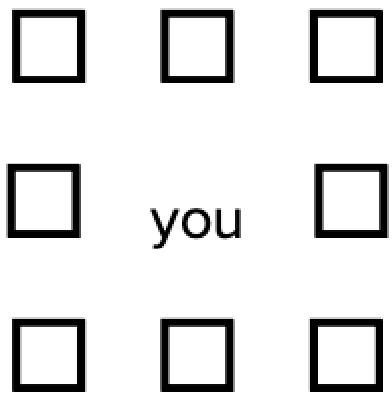

Figure 6.11 The figure in the questionnaire where audience reported acquaintances sitting around. The question is "The right image represents your seating position. The empty squares represent the people sitting around you. Please check the empty square if you know the person who is sitting in this position".

48 audience members in total attended the concert (female=28; male=19, 1 other or preferred not to say) between 19 and 83 years (mean =38.53, sd=15.09).

6.5.2 Results

Enhanced Frisson Occurrences

For all the 48 audience members, according to the questionnaire answers, 60% of them were familiar with the concept of frisson, 20% were familiar to a certain extent, and the remaining 20% audience members were not familiar with frisson. On a daily 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, while 40% felt frisson in relation to music “rarely” to “never”. In terms of frisson experiences in this concert, 40% of audience members reported that they felt more frisson during this concert than usual, while 8.3% felt less frisson. 6.25% reported feeling no frisson at all and the rest found it hard to answer. We found a strong correlation exists between the reported frequency of experiencing frisson normally and frisson during this concert ($R = .62$, $p < .001$).

For the Sharing and Non-sharing groups, there is a noticeable difference in the reported number of frisson occurrences during the concert. Although the Non-sharing group had reported feeling frisson in relation to music more often than the other group, the Non-sharing group did not experience more frisson during the concert than usual. 41.67% of the Non-sharing group reported that they had more frisson than usual, 54% of frisson-sharing group reported that they had more frisson than usual, the Sharing group had an 8.8% increase of frisson than usual, which suggests that the frisson-sharing mechanism does increase the number of frisson occurrences. This leads us to think that increasing frisson in a group which is originally less familiar with frisson is proof of the adequacy of the system. This result suggests our system worked as we expected for detecting and actuating neckbands to trigger frisson in a live concert scenario.

Enhanced Perceived Connectedness

In order to evaluate our hypothesis that the sense of connectedness in the Sharing group would be higher than in the Non-sharing group, we compared the subjective connectedness score between the two groups. However, these scores were found to correlate with the number of acquaintances sat adjacent to one another($R = .31$, $p = .03$). Among the Sharing group, the number of adjacent acquaintances (mean= 1.08, sd= 1.06) is 40% less ($p = .017$) than the Non-sharing group (mean=1.8 ,sd=1.32). The connectedness score in the Sharing group (mean= 3.54, sd=1.53) is 16% less ($p = .11$) than in the Non-sharing group(mean=4.17, sd=1.85). We cannot state a significant difference ($p = .11$) of the sense of connectedness within the two comparison groups, though we assume that the score should have a much stronger decrease in the Sharing group based on the significantly fewer ($p = .017$) acquaintances. Since it was an experiment in-the-wild study, the audience were allowed to pick their preferred seats, the groups were not balanced by personal relationships, and the Non-sharing group happened to know each other much better than the Sharing group. This highlights the importance of taking personal relationships between participants into consideration when evaluating connectedness.

6.6. Interpretation and Discussion

6.6.1 Trends of Group Dynamics

Because musical pieces in the three sessions are different, we interpreted what we found in the timeseries trends of EDA features referring to the characteristics of the musical pieces. Generally, we found the music parts with a strong dynamic, fast tempo, and contrasting changes could be reflected in increasing EDA Phasic values and SCR peaks counts.

In Session 1, the musical piece entered the theme in Variation I with weak and repetitive beats. It further developed into a more clear and brighter sentence from Variation II where EDA Phasic rised up (crescendo around 1). Similar changes could also be observed in the middle (fortissimo around 2) and the end of Variation III (forte around 2) when acoustic sounds and electric sound were integrated to create stronger dynamics and more abundant layers. In Session 2, the fourth variation in the Third Movement (around 18 minutes) became striking contrast to the preceding variation. This unexpected change and building up dynamics reflects in the peaks of average EDA Phasic (i). Similar patterns also occur in the sixth variation when it came to the treble and brilliant episode (ii).

In Session 3, when the dynamic gets stronger and the tempo gets faster in the middle part of Nocturne Op.27 No.1 (a and b), an abrupt change of EDA Phasic can be observed. The sudden surge also occurs in the last piece (No.24) of Chopin Preludes Op.28 [15] (c) when it arrived at the crescendo, fortissimo and brilliant sentence[16].

6.6.2 Physiological Entrainment Comparison between Sharing and Non-sharing Groups

The results from DTW indicated the entrainment of EDA tonic component in the Sharing group was lower than that in the Non-sharing group in the last musical session. While the entrainment of EDA Phasic in the Sharing group was higher than that in the Non-sharing group in the first two musical sessions. EDA tonic component indicates the slow change of skin conductance levels while the EDA Phasic reflects the quick and prompt change of skin conductance response [57, 60, 61]. Although wearing the device did not lead to more entrained feelings in the long-term trend. The thermal feedback did trigger much more entrained physiological reactions in the short-term and sudden arousal– which could be similarly interpreted as frisson occurrence. This result supported the feasibility of adopting a real-time frisson detection model and thermal feedback system to share this ambiguous and aesthetic feelings as a novel method to augment group dynamics by creating entrained experience. As proved by Tschacher et al. [138] and Gashi et al. [67], skin conductance entrainment could be associated with perceived engagement and appreciation during live events. Our results suggest this method of sharing frisson could be a potential way to enhance music appreciation.

Moreover, it is also worth discussing why no significant difference in physiological entrainment exists in Session 3. According to the project director, the musical piece in Session 3 is supposed to be less familiar but with more dramatic changes in musical elements than the first two sessions. Therefore, we assume that the musical piece in Session 3 could be so powerful in terms of creating frisson

15 https://www.youtube.com/watch?v=QHcEH2Rliko

16 https://en.wikipedia.org/wiki/Dynamics_(music)

events at certain timings that audience members could have similar physiological experience even without external triggering. If so, this might suggest the idea of detecting and sharing frisson could enhance the music appreciation experience especially when the music is more niche. However, there are still concerns about whether artificially generating entrained feelings could enhance perceived connectedness according to the subjective feedback from the audience group. Although we proved a correlation between reported connectedness and the number of adjacent acquaintances, we assume there is also a need to consider the risk of interrupting the natural music appreciation process.

6.7. Conclusion

This Chapter investigates the methodology of detecting and influencing individuals' physiological experience during in-the-wild group events. We developed a real-time frisson detection model with an average accuracy of 85.78% and described the detailed procedure of collecting labeled data, training, and evaluating the detection model. We implemented the frisson detection model in a real-life biofeedback system which could trigger frisson through a smart neckband via thermal feedback. By conducting the offline analysis on the collected dataset (the dataset is available for researchers under this link: https://osf.io/rzpn3/), we mapped EDA features' trends along the music development to understand the group dynamics. Our findings suggested the collective physiology could reflect certain characteristics of the music (e.g. dynamic and tempo), which matches the research insights we acquired in the previous chapters. We further applied DTW to quantify the physiological entrainment within groups. Compared with Non-sharing group, Sharing group experienced more physiological entrainment. Moreover, we also collected subjective feedback from two audience groups. The audience in the Sharing group overall presented a positive attitude towards this novel interaction and reported more frequent frisson experiences.

Our findings suggested the approach of recognizing and sharing frisson could be a potential way to augment music appreciation in groups. This approach could be applied to other ubiquitous sensing contexts to share physiological experience not only in co-located conditions but also in remote or even virtual conditions. However, it might bring possible risks, such as interfering one's natural physiological process (e.g. naturally occurring frisson) with triggered sensations (e.g. synthetic frisson). Iterations over design elements and analysis parameters in several controlled environments might help researchers and practitioners better prepare for the possible risks and challenges during in-the-wild implementations.

Chapter 7
Discussion and Implications

This dissertation explains a data analysis framework that could be integrated into biofeedback systems and applied to understand group dynamics. This Chapter firstly summarizes the ethical considerations in the projects mentioned in previous chapters and further discusses the implications for future applications to understand and augment group dynamics.

7.1. Ethical Considerations

Each study has received approval from the ethics committee at Keio University and was conducted according to the ethics rules and regulations of Keio University.

We explained how to collect and use their physiological data with the customized device before delivering the consent form. For the projects that real audience members attended, we only collected data from those who volunteered to join as study participants. Participants who volunteered to join the study can terminate their participation in the study at any time. Each participant signed a consent form before each study started. The consent form described the aim of the study, experiment steps, potential risks, and data protection policy. We helped participants wear the devices and confirmed they were wearing the device correctly and comfortably.

The data processing of each study is carried out in accordance with the data protection provisions of the General Data Protection Regulation (GDPR) and Keio University. The data collected can only be used exclusively for the purposes described in the consent form. Each participant was assigned identifiers in the dataset and data collected were only used anonymously. Personal information collected from demographic questionnaires and consent forms cannot be linked to the participant's identity.

7.2. Data Insights

Analysis could be considered as the process of transforming data into insights. This section discusses findings about extracting features from raw physiological data and extending individual physiological responses to group level reactions.

7.2.1 From Raw Data to Explainable Features

The projects we described in this thesis adopted BVP and EDA which are recorded continuously under a fixed sampling rate (see Table 7.1).

Preprocessing

Preprocessing is one essential step before feature extraction. There are several main methods to clean the physiological data in my analysis process:

- Inspect data types and units and keep consistency. Transform timestamps to certain time units if necessary.
- Remove the data before the event/activity started and after the event/activity ended. Because there could be random movement and noisy data especially when researchers were helping participants wear the devices.
- Remove or interpolate noisy data with the help of other recorded data such as accelerometer data and observational data (the analysis in Chapter 5 adopted this method).
- Add low pass/High pass/Band pass filters to resample the raw data. Low pass filters were applied to the analysis of the dataset described in this thesis to smooth the raw data. Table 7.1 summarizes the cutoff sampling rate adopted in the analysis.
- Inspect the waveform in BVP data to check whether regular and obvious peaks exist. Enhance peaks if necessary (the analysis in Chapter 4 adopted this method).

- Normalize or standardize feature data to reduce the individual difference before aggregating and comparing individual data. HRV features could be normalized by being divided by mean RR intervals besides common methods (e.g. MinMax scaler) [114–116].

Table 7.1 Raw physiological data and features used in the four projects for evaluations. In the column of filter parameters, the cutoff frequencies in the low-pass filters were summarized (more detailed information could be found in each chapter's analysis process).

Project	**Signal and Sampling Rate**	**Filter Parameters**	**Features Used**
Social Game Players Group (Chapter 4)	BVP (200Hz), EDA (10Hz)	BVP (4Hz), EDA (0.5Hz)	LF/HF ratio, RMSSD EDA Tonic, SCR peaks
Online Learning Students Group (Chapter 4)	BVP (100Hz)	BVP (4Hz)	pNN50, RMSSD
Contemporary Dance Audience Group (Chapter 5)	BVP (50Hz), EDA (4.545Hz)	BVP (3.5Hz), EDA (0.01Hz)	LF/HF ratio, pNN50, EDA difference, EDA extrema
Piano Concert Audience Group (Chapter 6)	BVP (50Hz), EDA (4.545Hz)	BVP (4Hz), EDA (0.5Hz)	pNN50, pNN20 MeanNN, EDA Tonic EDA Tonic difference EDA Phasic

The above methods could be applied at the same time. However, paying attention to the orders could enable better preprocessing performance (e.g. adjusting timestamp before removing unusable data). Moreover, some methods need to be adapted to real-time analysis according to the way of receiving data. For example, since data type should be consistent during data streaming, the size of the data chunk should be inspected before feature extraction. Because it is usually required around one to two minutes of BVP data to calculate stable HRV features.

Choices of HRV and EDA Features

To extract and select suitable features, we found the following aspects should be considered:

- Neuroscience proof: Which branch's activation of ANS (PSNS and SNS) is the main focus for the observation? Which features are more representative of that branch's activation?
- Time requirement: How long is the minimal time required for recording to calculate valid features? How long will it take to extract certain features (especially in real-time analysis)?
- Sensitivity: How sensitive/stable are the features? Which features could be more sensitive to catch rapid and short-term changes? Which features could. be more stable and robust to reflect long-term changes?

Overall, EDA features are considered as directly linked to SNS activation and more sensitive to short sudden elicitation compared to HRV-based features [61]. Based on prior works and our findings, EDA Phasic and EDA difference are more related to short-term arousal compared with EDA Tonic. As a reflection of the changes in EDA, EDA difference is generally quite discriminable and well-suited to gauging people's reactions in real-time, particularly during short temporal moments like direct interactions or shock-effects [7]. The changes in the tonic component of EDA (also known as skin conductance level – SCL) tend to be more stable, which could be related to more fundamental arousal dynamics. The phasic component of EDA (also known as skin conductance response – SCR) especially peaks in SCR (also known as SCR peaks) reflects short-term and event-related aroused responses. Therefore, aggregated SCR peaks and EDA extrema, are more suitable to identify key events and moments when the majority of the group members experience increased arousal. In the real-time analysis, those aggregated features might be used as a threshold to trigger biofeedback or present stable trend based on the major group dynamics in more formal conditions as similarly adopted in the prior work by Hassib et al [11]. However, the calculation relies on comparing to a global average and therefore a large chunk of data is required.

HRV features could be related to either SNS or PSNS activation or even both [52]. In the features selected in our four projects, pNN50 and RMSSD have been proven to be more related to PSNS while LF/HF ratio could be an indicator of the balance between SNS and PSNS activity [52]. As a frequency domain feature, LF/HF ratio needs more recording time to be calculated as stable values than the other time domain features. However, when more frequent fluctuations are expected especially to generate more vivid biofeedback in real-time, LF/HF ratio could be a good choice to reflect moment-by-moment experience. Specifically, we found LF/HF ratio could be a feature used in aesthetic or musical applications, which is also supported by some of the prior works [98, 99]. However, the interpretation of the LF/HF ratio could be controversial due to the complex nature of LF power [96]. Therefore, we would not suggest LF/HF ratio feature as an essential indicator when conducting offline analysis. pNN50 and RMSSD are under the control of PSNS. Especially, pNN50, calculated as the difference between adjacent heart periods, is nominally independent of resting HR [143]. This makes PNN50 relatively representative of the PSNS associated reactions such as relaxation and sustained attention [47, 92, 93, 121]. Our findings from contemporary dance audience groups over three performances suggested pNN50 could be a robust indicator of relaxation in group dynamics. We also suggest using pNN50 in real-time analysis to generate biofeedback when there are obvious contextual affect changes, such as tension and relief, or conflict and reconciliation. RMSSD reflects the HF power's variation in HR and could be extracted from short-term recordings [114, 144]. As an indicator for cognitive load [145], RMSSD could perform well to quantify entrainment between group members both in offline analysis and real-time analysis considering the low requirement for recording time.

7.2.2 From Individual Response to Group Dynamics

Following the concept of entrainment and evaluations over four projects, we found several analysis methods to infer collective reactions from individual responses or dyadic interactions.

Trend

The average value, together with the variance, among individuals' normalized feature data could be observed as a trend in the group dynamics. By inspecting the trend and variance, we could find the collective reactions as well as the timings when group members' experience diverge. Moreover, the trend could be more effective to reflect the group dynamics when there is a relatively explicit structure or plan for the event. Because the external stimuli could elicit similar and coordinated feelings as a trigger for induction entrainment in the group [138]. Therefore, we could interpret the trend together with specific structures, key moments, and other particular designs to understand the group dynamics from multiple perspectives. For example, Chapter 4 describes DTW results as a similarity metric that were plotted into trends. Compared with the workshop agenda, we found more similar cardiovascular reactions (reflected in pNN50) occurred among the learners with biofeedback after they started group discussions. Chapter 5 reports the results by comparing the trends of the audience's HRV and EDA features with the choreography where we found the pNN50's rising matched the choreography from relaxation to tension. Chapter 6 presents the correlations between the trends of collective EDA Phasic and music pieces' development.

Aggregated Value

Individual's normalized feature data could be aggregated in the following fashions:

- Count the number of group members who have similar experience in certain time range (e.g. aggregated SCR peaks counts in Chapter 4 and experience EDA extrema in Chapter 5).
- Categorize individual feature data according to structured periods (e.g. choreographic sections in Chapter 5).

It is essential to normalize the feature data before aggregation. Otherwise, extreme values due to the individual difference could bias the aggregated metrics. Further statistical methods could be applied to compare the aggregated values. For example, in Chapter 5, we aggregated the timeseries of HRV and EDA features to produce statistics for each of the six main choreographic sections. A repeated measures ANOVA with a Greenhouse-Gessier correction was used to investigate the correlation and variance. We found significant difference occurred among sections especially in pNN50, which supports our assumption that pNN50 could reflect the transition from tension to relaxation between the first half and second half of the performance.

Similarity

Similar reactions could be firstly inspected through the trends together with the variations among each individual member. We further applied dynamic time warping (DTW) [70], as an established measurement to calculate entrainment in previous works [67,71,72], to help us understand group dynamics. Although we selected DTW for the analysis described in this thesis, other analysis methods, such as pearson correlation [73, 74], cross recurrence quantification analysis (CRQA) [37], wavelet coherence analysis [68, 75], and machine learning algorithms [76, 77] could also be applied to quantify entrainment.

As a pairwise analysis method to quantify entrainment, DTW measures could be calculated for every pair within each group and further compared via statistical methods. For example, in Chapter 6, we compared DTW measures between frisson Sharing and Non-sharing groups and found the entrainment of EDA Phasic in the Sharing group was higher than that in the Non-sharing group in the first and second musical sessions. Moreover, the mean of the DTW measures among all the pairwise combinations within the group could be considered as the entrainment measurement of the whole group. In Section 4.4, the trends of average DTW among group pairs were plotted to reflect the development of entrainment during workshops.

7.3. Apply the Framework in Practice-led Research

Evaluations and projects presented in this thesis are primarily practice-led, where research methods, contexts, and outputs involve a significant focus on creative practice [27, 128, 129]. Different from traditional academic research, practice-led research could be more process-driven than goal-oriented [129]. This could lead to more challenges to collaborate with other team members with different backgrounds and interpret the data results with more uncontrollable variables. On the other hand, practice-led research has provided us with more opportunities to understand live group dynamics in a more natural environment. The design of event structure, especially the choreography and dance movements, could be relatively more controlled settings to collect in-the-wild dataset [146]. Moreover, practice-led research has allowed us to validate both real-time analysis and offline analysis, enabling us to contribute novel biofeedback systems. On the other hand, findings from the offline analysis could also help us evaluate the biofeedback system afterwards. This section summarizes the implications for interdisciplinary collaborations and biofeedback design when applying the framework.

7.3.1 Implications for Interdisciplinary Collaboration

Balance around Goals

The goals of researchers, practitioners, and artists can be very different even in the same project, so it is important for the project's success to uncover shared goals [29]. One shared goal in our project described in Chapter 5 was to explore and enhance the invisible link between the dancers and the audience through performance. Although in this work we prioritized artistic values such as the consistency of the theme and the immersive experience of the audience, the choreographer worked closely with researchers to include performance sections that were explicitly designed to trigger clear emotional changes - changes that prior evidence suggested would trigger physiological responses. Also in the project described in Section 4.4, the shared goal is to enhance engagement during online learning. However, we found the wrist-band type sensing device might interfere with students' learning behaviors (e.g. typing and taking notes) while researchers want to record physiological data for offline analysis. Around this shared goal, we revised the prototype to the ear-based device to reduce the risk of distractions.

Negotiate through Practices

Regular meet-ups at each stage of the co-design process are essential, especially during projects involving a large number of team members with different backgrounds (projects described in Chapter 5 and Chapter 6). The project described in Chapter 5 is a long-term project lasting approximately three years, which required researchers to attend major rehearsals, observe the stage conditions, and test prototypes on the spot. In this project, the artistic director connected the dance team and research team and led the negotiations by conveying expected choreographic elements and showing sensor feedback samples. Considering the pivotal role of music in the work, the dancers were also given access to samples of audio feedback as it was developed. The two teams met regularly and organized workshops to make and revise design choices. The following summarized the schedule of three workshops and key meetups when we were designing the feedback after the test performance:

The following summarized the schedule of three workshops and key meetups when we were designing the feedback after the test performance

- Meetup (2019. November): Shared feedback from the test performance and discussed the sensing feedback design schedule.
- Workshop (2019. December): The iteration started from a workshop where the choreography and the main piece – Bolero were introduced to the researcher team. Meanwhile, the researcher team prepared the hardware tryouts to help the dancers generate intuitions about physiological sensing.
- Workshop (2021. January): Discussed the performance choreographic sections' plan.
- Meetup (2021. January): Recorded sound elements used in the feedback loop. Rehearsed and adjusted the composed music with sound feedback.
- Workshop (2021. January): Mixed the sound, music, visual, and choreography together.
- Meetup (2021. March): Showed full performance together with all technological set-ups to the Session house staff.

This process not only contributed to a successful performance integrated with biofeedback but also enabled researchers to gain more perspectives to analyze the audience experience.

Share Research Insights

It is important that any findings and insights uncovered by the research team are regularly shared with other team members for both adjusting real-time analysis and conducting offline analysis. In the projects described in Chapter 6, when we were developing a real-time frisson detection model, we found the reported frisson events might be insufficient for the training based on the subjective feedback collected. After sharing this research insight with the director and engineering team, we first decided to play the music that proved to be effective in provoking chills [140] to the participants when we collected labeled data. We also used thermal feedback in the customized neckband to trigger more frisson events. All of these steps supported the development of the real-time frisson detection model and the final implementation of the frisson sharing system. In the projects described in Chapter 5, following the performances, we shared a version of Figure 5.5 with the dance team. Revealing the mapping of physiology and choreography in this way helped provide a fertile ground for further discussion. During the discussion, dancers matched scenes to changes in the graphs and shared their feelings, experiences, and audience comments around those specific moments. Some of the dancers mentioned they had trouble understanding the data visualization during the performance. However, looking back on the data afterwards provided them more time and space to consider the effects. They even further reflected on how to improvise while referring to the feedback loop system and contributed valuable insights to the interpretation of physiological data. This process was crucial to our co-creation project and helped us plan the way for future collaboration.

7.3.2 Implications for Biofeedback Design

In this thesis, we applied real-time analysis to generate biofeedback in the form of monitoring and reflecting physiological data mainly by visualizations. We further explored detecting and sharing certain physiological experience through thermal feedback. Results from offline analysis proved the increasing physiological entrainment for groups with biofeedback provided. This section will further discuss the influence and implications for implementing biofeedback to augment group interactions based on subjective feedback we received.

Following the concept of augmenting entrainment, we focus on investigating the feedback about connectedness and the sense of unity. Among all the projects, we found the implementation of biofeedback could enhance the entrained feelings for varied reasons. Table 7.2 summarizes the perceived connectedness types and advice for biofeedback design extracted from participants' subjective feedback:

Connected Counterparts	**Reported Feelings**	**Implications**
Event	Able to influence the event directly/indirectly	Clear mapping between people's physiological data and feedback
System	The system could promptly and correctly reflect their physiological states	Reduce time lag, Enhance perceived agency and control, Increase sensing accuracy
Other Members	Have similar feedback and be aware of contributing to the same activity	Provide collective goal or feedback

Table 7.2 Implications for biofeedback to enhance the sense of connectedness according to participants' feedback over four projects.

As suggested by Khut et al., fluent interactions, cognition, and expressive experiences are essential elements to create interactive experiences in biofeedback systems [22]:

- Fluent interactions: ability of interacting in an effortless and engaging way.
- Cognition: process of understanding the link between physiological data and biofeedback.
- Expressive experiences: moments of self-identification during the biofeedback.

Together with our collected subjective feedback, we suggested following implications for biofeedback design to augment group interactions.

Awareness of Individual's Experience and Agency

Before experiencing biofeedback in a collective manner, starting from an individual's biofeedback experience could help participants to understand how their own physiological data could influence the biofeedback. Simple interaction or biofeedback design techniques (e.g. customizing the appearance of biofeedback elements [22]) could be implemented for participants to catch this agency. For example, although we did not describe the biofeedback setup of the project in Chapter 5, we designed biofeedback starting from reflecting individual's physiological data before generating collective biofeedback (see Table 5.1 and Figure 5.2). However, in the project described in Section 4.3, the line charts generated from the average values of data contributors' HRV features failed to take care for the individual's agency. Although we decided to use the aggregated value directly due to privacy concerns, we noticed there could be alternative methods to reinforce the link and agency. For example, introducing error bars or standard deviations when generating line charts could reflect the ability to influence biofeedback to some degree.

Timings of Enhancing Entrained Experience

Based on the subjective feedback, we found participants may appreciate sharing physiological reactions within the group more at certain timings while sometimes understanding others' feelings may not help the overall experience. For example, we found some of the audience members attending the contemporary dance performance whether the sense of unity was necessary when enjoying this aesthetic performance (see in Chapter 5). Moreover, we also worried whether manipulating audience's physiological reactions by providing thermal feedback would affect the natural frisson feelings as explored in Chapter 6. However, considering the positive feedback from participants about experiencing this novel interaction, we still assume it is worthwhile to continue exploring biofeedback as a new way to appreciate artistic events. This would require HCI researchers and practitioners to investigate the various needs of different people and provide multiple choices to trigger biofeedback at appropriate timings.

Chapter 8
Conclusion and Future Directions

This Chapter presents an overview of the dissertation and reviews research questions put forward in Chapter 1. The contributions of this research are highlighted, followed by future directions derived from this work.

8.1. Dissertation Overview

This research work presents a practical framework for physiological data analysis to understand and augment group dynamics. Following the concept of entrainment, as related to the presence of similar reactions among group members [37], we introduce a real-time analysis component to trigger and an offline analysis to investigate similar physiological experience at group level. The frameworks was explained and evaluated in practice-based research projects from initial exploration in small group interactions to in-the-wild large scale group events.

Chapter 4 presents methods to quantify group dynamics by investigating trends and similarity between physiological data followed by initial explorations for sharing physiological experience over visualizations. Chapter 5 further extends the exploration to understand large-scale group dynamics during in-the-wild group events and provides a feasible mapping between collective physiological trends and notable moments with choreography. Chapter 6 presents a complete flow of developing real-time algorithms to generate biofeedback at real-life concerts, evaluating the effect of the biofeedback system, and mapping group dynamics to musical piece's unfolding.

Besides demonstrations of the analysis process and summaries of the findings, this dissertation discusses how to apply the framework in practice-led research. In Chapter 7, we firstly discussed about extracting explainable features and calculating group level metrics for analysis. We further contributed implications for interdisciplinary collaboration and biofeedback design. We assume this research works could help researchers and practitioners in the HCI field to novel measurements and experiences in terms of augmenting group dynamics.

8.2. Research Questions Review

This section presents a review of the research questions and corresponding discussions based on the evaluations:

Research Question 1: How can we use the concept of entrainment to improve understanding of group dynamics by physiological data? We probed the methodologies of analyzing physiological data focusing on the trends of group dynamics and similar reactions among group members. The specific research questions are answered as follows:

(a) How can the proposed offline analysis be used to quantify group dynamics beyond individual subjective experience? Based on the offline analysis we conducted and the research insights we acquired, we summarized essential steps as reported in Chapter 7.

Firstly, it is essential to clean physiological data before further feature extraction and analysis. Preprocessing could be conducted by removing noisy data and applying filters etc. Explainable features could be extracted and selected by understanding the characteristics of the sensing modality. Take HRV features and EDA features we have explored for example. pNN50 has been proved to be closely related to PSNS activation, thus more relaxed feeling could be reflected in increasing pNN50. EDA features, such as peaks, are more related to SNS activation, thus more peaks could indicate more emotional arousal. Further, we also found several analysis methods to achieve an understanding of group experience beyond individual physiological reactions. For example, we could calculate average values and variance among group members or aggregate numbers of people whose physiological data show similar patterns within certain time periods. Finally, using the concept of entrainment could help us interpret the results. One perspective to interpret the result could be exploring the possible entrainment triggered by emotional empathy within the group. Another perspective is investigating the potential entrainment that could be triggered by structural factors such as the choreography of the performance and the rhythms of the musical pieces. Linking the collective physiology to the development of the group event may provide a holistic view of the in-the-wild group experience. In summary, the concept of entrainment could improve understanding of group dynamics. Specifically, this concept could suggest feasible analysis methods to extract collective physiology and provide interpretation perspectives to understand group experience.

(b) Which aspects can research insights acquired in offline analysis imply real-time analysis in biofeedback systems?

Reflecting on the applications and evaluations of the described projects, we found research insights acquired in the offline analysis could provide meaningful references for real-time analysis in biofeedback systems. One important valuable aspect is regarding feature choices as discussed in Chapter 7. From offline analysis, we investigated the features' robustness, sensitivity, and computing cost, which are key aspects of a real-time biofeedback system. Moreover, through careful calculation and analysis, we found the links between certain features and psycho-physiological states (e.g. pNN50 is suitable to reflect the group's relaxation level). Additionally, most of the analysis methods we explored in the offline analysis could actually be conducted on a relatively short window of the dataset such as calculating mean values and conducting predictions. In summary, research insights acquired from the offline analysis could imply the choices of features and play a pivotal role when adjusting parameters in real-time analysis algorithms.

Research Question 2: How can we use the concept of entrainment to improve augmenting group dynamics by physiological data?

We explored the methodologies of developing real-time analysis algorithms for sensor-based interactions ranging from reflecting, detecting, and sharing physiological experience within groups. The specific research questions are answered as follows:

(a) How can the proposed framework for physiological data analysis be applied to real-life biofeedback systems?

Based on our practices where we implemented real-time analysis to biofeedback systems, we extracted a common workflow. The workflow consists of receiving physiological data from group members, calculating and analyzing data via python scripts running at servers, and transmitting results to influence the outputs as feedback. In the python script, steps including preprocessing, feature extraction, analysis, and aggregation are written (see Appendix C for an example of python script). Parameters, such as window size and feature types, could be adjusted flexibly according to the requirements. Feedback could be embedded in either existing platforms (e.g. Plotly dashboard described in Chapter 4) or self-built biofeedback systems (e.g. frisson sharing systems described in Chapter 6).

(b) What effects do the biofeedback systems embedded with the proposed real-time analysis bring to group interactions?

From the offline analysis on the collected physiological dataset, we found participants who experienced the biofeedback systems tend to have more physiological entrainment. This might suggest biofeedback systems could augment group interaction by enhancing entrained feelings at the physiological level. As for perceived connectedness and entrainment collected from subjective feedback, some participants mentioned their increasing sense of unity and feelings of being connected, while some participants did not. Moreover, we found possible risks of bringing more cognitive load and privacy concerns are worth considering when sharing the physiological experience with others.

Research Question 3: How to integrate the proposed framework with practical goals during interdisciplinary collaborations?

During the process of evaluation, we tested the practical framework for physiological data analysis in interdisciplinary collaborations (researchers and practitioners in the HCI field, professional artists, and experts with various domain knowledge). The ultimate goal of each project was usually creating a novel experience or installation with biofeedback implemented. While approaching the goal, we tested and revised the analysis algorithms for better performance. Offline analysis was conducted on data collected during the event to reflect live group dynamics and to evaluate the biofeedback system. More detailed implications for applying the framework to interdisciplinary collaborations were summarized in Chapter 7.

8.3. Limitations

First, although we tried to sense the live group in an unobtrusive way with wearable devices, it was still hard to totally remove the interference. We explored the potential of collecting data via ear-based devices in the project described in Chapter 4 (Section 4.4) to reduce the awareness of being sensed. However, ear-based devices might need more time for calibrations and adjustment because the head shape difference turned out to be more obvious than that of fingers.

We were not able to obtain valid data from people who were moving too much since both EDA and BVP are sensitive to movement artifacts. Especially when people were interacting with each other and the group event lasted for a long time. There are also alternative methods to analyze the dataset in terms of temporal component and rhythms [8, 147] or sub-group clustering [7] to further understand how entrainment took place.

The dataset for offline analysis in Chapter 5 was collected from a real performance where audience physiological data was used to trigger changes in staging elements. The existence of the feedback loop complicated the exploration and interpretation of the results. In Chapter 4 (Section 4.4), we tried dividing participants into control and experiment groups. However, some participants in the control group expressed disappointment, which was also reported by the audience group attending the test performance in the project described in Chapter 5. We consider this is also a challenging but valuable topic for further investigations and related research.

8.4. Future Directions

This dissertation describes how to analyze physiological data to understand group dynamics. We found several methods and interpretation perspectives to translate this type of abstract and raw data into explainable results. Yet we must admit it is not sufficient to rely on the physiological data alone to quantify group reactions. Moreover, the valence dimension is relatively hard to detect compared with the arousal dimension. In the future, mappings between valence levels and physiological reactions or even more diverse emotion models could be considered when investigating affective reactions.

As for biofeedback design based on detecting and analyzing affective reactions, finding a proper way to utilize physiological data with minimal distraction and intuitiveness for people to understand the meaning behind biofeedback is essential. Current feedback from participants expresses confusion when there was no explicit connection between biofeedback and their physiological data (e.g. the color became red when they felt stress). One direction could be developing and testing to find out a more clear linkage and informing those who will experience the biofeedback in advance. On the contrary, we would also expect more implicit but intuitive interactions where people could have minimal feelings of being monitored.

Moreover, in the analysis we presented in this thesis, we tend to interpret higher entrainment levels as more positive aspects in group dynamics (e.g. increasing connectedness). Biofeedback systems embedded with real-time analysis algorithms are also expected to increase entrained experience. However, it is worth further exploring the correlation between entrainment and group dynamics because entrainment might not always be related to positive experience [73]. It is especially important to test and discuss when should we enhance entrainment and when should we alleviate entrainment when designing for biofeedback.

Publications

This chapter summarizes international peer-reviewed publications divided into journal publications and conference publications.

Peer-Reviewed Journal Publications

1. **Jiawen Han**, George Chernyshov, Moe Sugawa, Dingding Zheng, Danny Hynds, Taichi Furukawa, Marcelo Padovani, Karola Marky, Kouta Minamizawa, Jamie A Ward, Kai Kunze. "Linking Audience Physiology to Choreography." ACM Transactions on Computer-Human Interaction (2021).

2. Yan He, George Chernyshov, **Jiawen Han**, Dingding Zheng, Ragnar Thomsen, Danny Hynds, Muyu Liu et al. "Frisson Waves: Exploring Automatic Detection, Triggering and Sharing of Aesthetic Chills in Music Performances." Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 6, no. 3 (2022): 1-23.

Peer-Reviewed Conference Publications

1. **Jiawen Han**, Chi-Lan Yang, George Chernyshov, Zhuoqi Fu, Reiya Horii, Takuji Narumi, and Kai Kunze. "Exploring Collective Physiology Sharing as Social Cues to Support Engagement in Online Learning." In 20th International Conference on Mobile and Ubiquitous Multimedia, pp. 192-194. 2021.

2. Sugawa, Moe, 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, pp. 1-10. 2021.

3. Zhuoqi Fu, **Jiawen Han**, Dingding Zheng, Moe Sugawa, Taichi Furukawa, Chernyshov George, Hynds Danny et al. "Boiling Mind-A Dataset of Physiological Signals during an Exploratory Dance Performance." In Augmented Humans Conference 2021, pp. 301-303. 2021.

4. Yan He, George Chernyshov, Dingding Zheng, **Jiawen Han**, Ragnar Thomsen, Danny Hynds, Yuehui Yang, Yun Suen Pai, Kai Kunze, and Kouta Minamizawa. "Frisson Waves: Sharing Frisson to Create Collective Empathetic Experiences for Music Performances." In SIGGRAPH Asia 2021 Emerging Technologies, pp. 1-2. 2021.

5. Kanyu Chen, **Jiawen Han**, George Chernyshov, Christopher Kim, Ismael Rasa, and Kai Kunze. 2021. "Affective Umbrella – Towards a Novel Sensor Integrated Multimedia Platform Using Electrodermal and Heart Activity in an Umbrella Handle." In 20th International Conference on Mobile and Ubiquitous Multimedia (MUM 2021). Association for Computing Machinery, New York, NY, USA, 208–210.

6. Christopher Changmok Kim, **Jiawen Han**, Dingding Zheng, George Chernyshov, and Kai Kunze. 2021. "Using Smart Eyewear to Sense Electrodermal Activity While Reading." In Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers (UbiComp '21). Association for Computing Machinery, New York, NY, USA, 472–475.

7. Xiaru Meng, **Jiawen Han**, George Chernyshov, Kirill Ragozin, and Kai Kunze. 2022. "ThermalDrive - Towards Situation Awareness over Thermal Feedback in Automated Driving Scenarios." In 27th International Conference on Intelligent User Interfaces (IUI '22 Companion). Association for Computing Machinery, New York, NY, USA, 101–104.

8. Jie Bao, **Jiawen Han**, Akira Kato, and Kai Kunze. 2020. "Sleepy watch: towards predicting daytime sleepiness based on body temperature." In Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers (UbiComp-ISWC '20). Association for Computing Machinery, New York, NY, USA, 9–12.

9. **Jiawen Han**, George Chernyshov, Dingding Zheng, Peizhong Gao, Takuji Narumi, Katrin Wolf, and Kai Kunze. 2019. "Sentiment Pen: Recognizing Emotional Context Based on Handwriting Features." In Proceedings of the 10th Augmented Human International Conference 2019 (AH2019). Association for Computing Machinery, New York, NY, USA, Article 24, 1–8.

References

[1] Rosalind W Picard. *Affective computing*. MIT press, 2000.

[2] Walter Bradford Cannon. *Bodily changes in pain, hunger, fear and rage: An account of recent researches into the function of emotional excitement*. D. Appleton, 1922.

[3] Sylvia D Kreibig. Autonomic nervous system activity in emotion: A review. *Biological psychology*, 84(3):394–421, 2010.

[4] Richard V Palumbo, Marisa E Marraccini, Lisa L Weyandt, Oliver Wilder-Smith, Heather A McGee, Siwei Liu, and Matthew S Goodwin. Interpersonal autonomic physiology: A systematic review of the literature. *Personality and Social Psychology Review*, 21(2):99–141, 2017.

[5] Celine Latulipe, Erin A Carroll, and Danielle Lottridge. Love, hate, arousal,and engagement:exploring audience responses to performing arts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1845–1854, 2011.

[6] Fernando Silveira, Brian Eriksson, Anmol Sheth, and Adam Sheppard. Predicting audience responses to movie content from electro-dermal activity signals. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 707–716, 2013.

[7] Chen Wang, Erik N Geelhoed, Phil P Stenton, and Pablo Cesar. Sensing a live audience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1909–1912, 2014.

[8] Asaf Bachrach, Yann Fontbonne, Coline Joufflineau, and José Luis Ulloa. Audience entrainment during live contemporary dance performance:

physiological and cognitive measures. *Frontiers in Human Neuroscience*, 9:179, 2015. URL: `https://www.frontiersin.org/article/10.3389/fnhum.2015.00179`, `doi:10.3389/fnhum.2015.00179`.

[9] Haruka Shoda, Mayumi Adachi, and Tomohiro Umeda. How live performance moves the human heart. *PloS one*, 11(4):e0154322, 2016.

[10] Staci Vicary, Matthias Sperling, Jorina von Zimmermann, Daniel C. Richardson, and Guido Orgs. Joint action aesthetics. *PLOS ONE*, 12(7):1–21, 07 2017. URL: `https://doi.org/10.1371/journal.pone.0180101`, `doi:10.1371/journal.pone.0180101`.

[11] Mariam Hassib, Stefan Schneegass, Philipp Eiglsperger, Niels Henze, Albrecht Schmidt, and Florian Alt. Engagemeter: A system for implicit audience engagement sensing using electroencephalography. In *Proceedings of the 2017 Chi conference on human factors in computing systems*, pages 5114–5119, 2017.

[12] Lida Theodorou, Patrick GT Healey, and Fabrizio Smeraldi. Engaging with contemporary dance: What can body movements tell us about audience responses? *Frontiers in psychology*, 10:71, 2019.

[13] Dana Swarbrick, Dan Bosnyak, Steven R Livingstone, Jotthi Bansal, Susan Marsh-Rollo, Matthew H Woolhouse, and Laurel J Trainor. How live music moves us: head movement differences in audiences to live versus recorded music. *Frontiers in psychology*, 9:2682, 2019.

[14] E. Gedik, L. Cabrera-Quiros, C. Martella, G. Englebienne, and H. Hung. Towards analyzing and predicting the experience of live performances with wearable sensing. *IEEE Transactions on Affective Computing*, 12(1):269–276, 2021. `doi:10.1109/TAFFC.2018.2875987`.

[15] Alessandro Vinciarelli, Hugues Salamin, and Maja Pantic. Social signal processing: Understanding social interactions through nonverbal behavior analysis. In *2009 IEEE computer society conference on computer vision and pattern recognition workshops*, pages 42–49. IEEE, 2009.

[16] Tanya L Chartrand and John A Bargh. The chameleon effect: the perception–behavior link and social interaction. *Journal of personality and social psychology*, 76(6):893, 1999.

[17] Arkady Pikovsky, Michael Rosenblum, Jürgen Kurths, and A Synchronization. A universal concept in nonlinear sciences. *Self*, 2:3, 2001.

[18] Martin Clayton. What is entrainment? definition and applications in musical research. 2012.

[19] William S Condon and Louis W Sander. Synchrony demonstrated between movements of the neonate and adult speech. *Child development*, pages 456–462, 1974.

[20] Carlos Cornejo, Zamara Cuadros, Ricardo Morales, and Javiera Paredes. Interpersonal coordination: methods, achievements, and challenges. *Frontiers in psychology*, 8:1685, 2017.

[21] Adam J Strang, Gregory J Funke, Sheldon M Russell, Allen W Dukes, and Matthew S Middendorf. Physio-behavioral coupling in a cooperative team task: contributors and relations. *Journal of Experimental Psychology: Human Perception and Performance*, 40(1):145, 2014.

[22] George Khut. Development and evaluation of participant-centred biofeedback artworks. 2006.

[23] Petr Slovák, Joris Janssen, and Geraldine Fitzpatrick. Understanding heart rate sharing: towards unpacking physiosocial space. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 859–868, 2012.

[24] R Michael Winters, Bruce N Walker, and Grace Leslie. Can you hear my heartbeat?: hearing an expressive biosignal elicits empathy. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–11, 2021.

[25] Jaime Snyder, Mark Matthews, Jacqueline Chien, Pamara F Chang, Emily Sun, Saeed Abdullah, and Geri Gay. Moodlight: Exploring personal and

social implications of ambient display of biosensor data. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, pages 143–153, 2015.

[26] Kristina Höök, Martin P Jonsson, Anna Ståhl, and Johanna Mercurio. Somaesthetic appreciation design. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 3131–3142, 2016.

[27] Steve Benford, Chris Greenhalgh, Andy Crabtree, Martin Flintham, Brendan Walker, Joe Marshall, Boriana Koleva, Stefan Rennick Egglestone, Gabriella Giannachi, Matt Adams, et al. Performance-led research in the wild. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 20(3):1–22, 2013.

[28] Graeme Sullivan. Research acts in art practice. *Studies in art Education*, 48(1):19–35, 2006.

[29] Celine Latulipe, Erin A Carroll, and Danielle Lottridge. Evaluating longitudinal projects combining technology with temporal arts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1835–1844, 2011.

[30] Maja Pantic, Alex Pentland, Anton Nijholt, and Thomas S Huang. Human computing and machine understanding of human behavior: A survey. In *Artifical intelligence for human computing*, pages 47–71. Springer, 2007.

[31] Douglas C Engelbart. Augmenting human intellect: A conceptual framework. *Menlo Park, CA*, 1962.

[32] Norbert Wiener. *Cybernetics or Control and Communication in the Animal and the Machine*. MIT press, 2019.

[33] Florian Floyd Mueller, Pedro Lopes, Paul Strohmeier, Wendy Ju, Caitlyn Seim, Martin Weigel, Suranga Nanayakkara, Marianna Obrist, Zhuying Li, Joseph Delfa, et al. Next steps for human-computer integration. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2020.

[34] Andy Clark. Natural-born cyborgs? In *International Conference on Cognitive Technology*, pages 17–24. Springer, 2001.

[35] Donna Haraway. A cyborg manifesto: Science, technology, and socialist-feminism in the late 20th century. In *The international handbook of virtual learning environments*, pages 117–158. Springer, 2006.

[36] Björn Hartmann, Leith Abdulla, Manas Mittal, and Scott R Klemmer. Authoring sensor-based interactions by demonstration with direct manipulation and pattern recognition. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 145–154, 2007.

[37] Martina Ardizzi, Marta Calbi, Simona Tavaglione, Maria Alessandra Umiltà, and Vittorio Gallese. Audience spontaneous entrainment during the collective enjoyment of live performances: physiological and behavioral measurements. *Scientific reports*, 10(1):1–12, 2020.

[38] Jiawen Han, Chi-Lan Yang, George Chernyshov, Zhuoqi Fu, Reiya Horii, Takuji Narumi, and Kai Kunze. Exploring collective physiology sharing as social cues to support engagement in online learning. In *20th International Conference on Mobile and Ubiquitous Multimedia*, pages 192–194, 2021.

[39] Jiawen Han, George Chernyshov, Moe Sugawa, Dingding Zheng, Danny Hynds, Taichi Furukawa, Marcelo Padovani, Kouta Minamizawa, Karola Marky, Jamie A Ward, et al. Linking audience physiology to choreography. *ACM Transactions on Computer-Human Interaction*, 2021.

[40] 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*, TEI '21, New York, NY, USA, 2021. Association for Computing Machinery. URL: `https://doi.org/10.1145/3430524.3440653`, `doi:10.1145/3430524.3440653`.

[41] 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.

[42] Yan He, George Chernyshov, Jiawen Han, Dingding Zheng, Ragnar Thomsen, Danny Hynds, Muyu Liu, Yuehui Yang, Yulan Ju, Yun Suen Pai, Kouta Minamizawa, Kai Kunze, and Jamie A. Ward. Frisson waves: Exploring automatic detection, triggering and sharing of aesthetic chills in music performances. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 6(3), sep 2022. URL: `https://doi.org/10.1145/3550324`, `doi:10.1145/3550324`.

[43] Yan He, George Chernyshov, Dingding Zheng, Jiawen Han, Ragnar Thomsen, Danny Hynds, Yuehui Yang, Yun Suen Pai, Kai Kunze, and Kouta Minamizawa. Frisson waves: Sharing frisson to create collective empathetic experiences for music performances. In *SIGGRAPH Asia 2021 Emerging Technologies*, pages 1–2. 2021.

[44] Robert W Levenson. Blood, sweat, and fears: The autonomic architecture of emotion. *Annals of the New York Academy of Sciences*, 1000(1):348–366, 2003.

[45] Bradley M Appelhans and Linda J Luecken. Heart rate variability as an index of regulated emotional responding. *Review of general psychology*, 10(3):229–240, 2006.

[46] Philip Schmidt, Attila Reiss, Robert Duerichen, and Kristof Van Laerhoven. Wearable affect and stress recognition: A review. *arXiv preprint arXiv:1811.08854*, 2018.

[47] Patricia J Bota, Chen Wang, Ana LN Fred, and Hugo Plácido Da Silva. A review, current challenges, and future possibilities on emotion recognition using machine learning and physiological signals. *IEEE Access*, 7:140990–141020, 2019.

[48] Robert W Levenson. The autonomic nervous system and emotion. *Emotion Review*, 6(2):100–112, 2014.

[49] Julian F Thayer and Esther Sternberg. Beyond heart rate variability: vagal regulation of allostatic systems. *Annals of the New York Academy of Sciences*, 1088(1):361–372, 2006.

[50] 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.

[51] Jos F Brosschot, Eduard Van Dijk, and Julian F Thayer. Daily worry is related to low heart rate variability during waking and the subsequent nocturnal sleep period. *International journal of psychophysiology*, 63(1):39–47, 2007.

[52] Fred Shaffer and JP Ginsberg. An overview of heart rate variability metrics and norms. *Frontiers in public health*, 5:258, 2017.

[53] Carl Gustav Jung. *Studies in word-association.* Heinemann, 1918.

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

[55] 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.

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

[57] Wolfram Boucsein. *Electrodermal activity.* Springer Science & Business Media, 2012.

[58] 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.

[59] Erin T Solovey, Marin Zec, Enrique Abdon Garcia Perez, Bryan Reimer, and Bruce Mehler. Classifying driver workload using physiological and driving performance data: two field studies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 4057–4066, 2014.

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

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

[62] William S Condon and William D Ogston. Sound film analysis of normal and pathological behavior patterns. *Journal of nervous and mental disease*, 1966.

[63] Melissa Ellamil, Joshua Berson, Jen Wong, Louis Buckley, and Daniel S Margulies. One in the dance: musical correlates of group synchrony in a real-world club environment. *PloS one*, 11(10):e0164783, 2016.

[64] Martin Lang, Daniel J Shaw, Paul Reddish, Sebastian Wallot, Panagiotis Mitkidis, and Dimitris Xygalatas. Lost in the rhythm: effects of rhythm on subsequent interpersonal coordination. *Cognitive Science*, 40(7):1797–1815, 2016.

[65] Richard C Schmidt and Michael J Richardson. Dynamics of interpersonal coordination. In *Coordination: Neural, behavioral and social dynamics*, pages 281–308. Springer, 2008.

[66] Liam Cross, Martine Turgeon, and Gray Atherton. How moving together binds us together: the social consequences of interpersonal entrainment and group processes. *Open Psychology*, 1(1):273–302, 2019.

[67] Shkurta Gashi, Elena Di Lascio, and Silvia Santini. Using unobtrusive wearable sensors to measure the physiological synchrony between presenters and audience members. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(1):1–19, 2019.

[68] Giorgio Quer, Joshal Daftari, and Ramesh R Rao. Heart rate wavelet coherence analysis to investigate group entrainment. *Pervasive and Mobile Computing*, 28:21–34, 2016.

[69] Peter E Keller, Giacomo Novembre, and Michael J Hove. Rhythm in joint action: psychological and neurophysiological mechanisms for real-time interpersonal coordination. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1658):20130394, 2014.

[70] Donald J Berndt and James Clifford. Using dynamic time warping to find patterns in time series. In *KDD workshop*, volume 10, pages 359–370. Seattle, WA, USA:, 1994.

[71] Theodoros Kostoulas, Guillaume Chanel, Michal Muszynski, Patrizia Lombardo, and Thierry Pun. Dynamic time warping of multimodal signals for detecting highlights in movies. In *Proceedings of the 1st Workshop on Modeling INTERPERsonal SynchrONy And infLuence*, pages 35–40, 2015.

[72] Eunice Jun, Daniel McDuff, and Mary Czerwinski. Circadian rhythms and physiological synchrony: Evidence of the impact of diversity on small group creativity. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–22, 2019.

[73] Maria Elide Vanutelli, Laura Gatti, Laura Angioletti, and Michela Balconi. Affective synchrony and autonomic coupling during cooperation: a hyperscanning study. *BioMed Research International*, 2017, 2017.

[74] Eetu Haataja, Jonna Malmberg, and Sanna Järvelä. Monitoring in collaborative learning: Co-occurrence of observed behavior and physiological synchrony explored. *Computers in Human Behavior*, 87:337–347, 2018.

[75] Jamie A Ward, Daniel Richardson, Guido Orgs, Kelly Hunter, and Antonia Hamilton. Sensing interpersonal synchrony between actors and autistic children in theatre using wrist-worn accelerometers. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers*, pages 148–155, 2018.

[76] Fahd Albinali, Matthew S Goodwin, and Stephen S Intille. Recognizing stereotypical motor movements in the laboratory and classroom: a case study with children on the autism spectrum. In *Proceedings of the 11th international conference on Ubiquitous computing*, pages 71–80, 2009.

[77] Emilie Delaherche, Mohamed Chetouani, Ammar Mahdhaoui, Catherine Saint-Georges, Sylvie Viaux, and David Cohen. Interpersonal synchrony: A survey of evaluation methods across disciplines. *IEEE Transactions on Affective Computing*, 3(3):349–365, 2012.

[78] Catherine J Stevens, Emery Schubert, Rua Haszard Morris, Matt Frear, Johnson Chen, Sue Healey, Colin Schoknecht, and Stephen Hansen. Cognition and the temporal arts: Investigating audience response to dance using pdas that record continuous data during live performance. *International Journal of Human-Computer Studies*, 67(9):800–813, 2009.

[79] Danilo Rodrigues, Emily Grenader, Fernando Nos, Marcel Dall'Agnol, Troels Hansen, and Nadir Weibel. Motiondraw: a tool for enhancing art and performance using kinect. In *CHI '13*, 2013. URL: `https://doi.org/10.1145/2468356.2468570`.

[80] Courtney Brown. Machine tango: An interactive tango dance performance. pages 565–569, 03 2019. `doi:10.1145/3294109.3301263`.

[81] Christopher Lindinger, Martina Mara, Klaus Obermaier, Roland Aigner, Roland Haring, and Veronika Pauser. The (st)age of participation: audience involvement in interactive performances. *Digital Creativity*, 24(2):119–129, 2013. URL: `https://doi.org/10.1080/14626268.2013.808966`, `doi:10.1080/14626268.2013.808966`.

[82] Asreen Rostami, Donald McMillan, Elena Márquez Segura, Chiara Rossito, and Louise Barkhuus. Bio-sensed and embodied participation in interactive performance. In *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction*, pages 197–208, 2017.

[83] Kristina Höök. Affective loop experiences–what are they? In *International Conference on Persuasive Technology*, pages 1–12. Springer, 2008.

[84] Brad Myers and Andrew Ko. The past, present and future of programming in hci. 2009.

[85] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[86] Anthony J Onwuegbuzie, Nancy L Leech, et al. Linking research questions to mixed methods data analysis procedures. *The qualitative report*, 11(3):474–498, 2006.

[87] Jason J Braithwaite, Derrick G Watson, Robert Jones, and Mickey Rowe. A guide for analysing electrodermal activity (eda) & skin conductance responses (scrs) for psychological experiments. *Psychophysiology*, 49(1):1017–1034, 2013.

[88] 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.

[89] Gheorghe Cernisov. *An unobtrusive wearable sensing platform for psychological and social dynamics (dissertation)*. PhD thesis, Keio University, 2020.

[90] Dingding Zheng. *An Interdisciplinary Psychometric Method for Psychophysiological Tracking of Social Interactions and Social Dynamics in "Real-World" Environments (dissertation)*. PhD thesis, Keio University, 2021.

[91] Kanyu Chen, Jiawen Han, George Chernyshov, Christopher Kim, Ismael Rasa, and Kai Kunze. Affective umbrella–towards a novel sensor integrated multimedia platform using electrodermal and heart activity in an umbrella handle. In *20th International Conference on Mobile and Ubiquitous Multimedia*, pages 208–210, 2021.

[92] Robert F Potter and Paul Bolls. *Psychophysiological measurement and meaning: Cognitive and emotional processing of media*. Routledge, 2012.

[93] Jacob B Holzman and David J Bridgett. Heart rate variability indices as biomarkers of top-down self-regulatory mechanisms: A meta-analytic review. *Neuroscience & biobehavioral reviews*, 74:233–255, 2017.

[94] Michael Trimmel. Relationship of heart rate variability (hrv) parameters including pnnxx with the subjective experience of stress, depression, well-being, and every-day trait moods (trim-t): A pilot study. *The Ergonomics Open Journal*, 8(1), 2015.

[95] Muneeb Imtiaz Ahmad, David A Robb, Ingo Keller, and Katrin Lohan. Towards a multimodal measure for physiological behaviours to estimate cognitive load. In *International Conference on Human-Computer Interaction*, pages 3–13. Springer, 2020.

[96] George E Billman. The lf/hf ratio does not accurately measure cardiac sympatho-vagal balance. *Frontiers in physiology*, 4:26, 2013.

[97] 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.

[98] Yuuki Ooishi, Hideo Mukai, Ken Watanabe, Suguru Kawato, and Makio Kashino. Increase in salivary oxytocin and decrease in salivary cortisol after listening to relaxing slow-tempo and exciting fast-tempo music. *PloS one*, 12(12):e0189075, 2017.

[99] Luciano Bernardi, Cesare Porta, and Peter Sleight. Cardiovascular, cerebrovascular, and respiratory changes induced by different types of music in musicians and non-musicians: the importance of silence. *Heart*, 92(4):445–452, 2006.

[100] 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`.

[101] Wolfgang Ambach and Matthias Gamer. Physiological measures in the detection of deception and concealed information. In *Detecting Concealed Information and Deception*, pages 3–33. Elsevier, 2018.

[102] Christopher T Conner and Nicholas M Baxter. Are you a werewolf? teaching symbolic interaction theory through game play. *Teaching Sociology*, 50(1):17–27, 2022.

[103] René F Kizilcec, Chris Piech, and Emily Schneider. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *Proceedings of the third international conference on learning analytics and knowledge*, pages 170–179, 2013.

[104] Saijing Zheng, Mary Beth Rosson, Patrick C Shih, and John M Carroll. Understanding student motivation, behaviors and perceptions in moocs. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, pages 1882–1895, 2015.

[105] Na Sun, Mary Beth Rosson, and John M Carroll. Where is community among online learners? identity, efficacy and personal ties. In *Proceedings of the 2018 chi conference on human factors in computing systems*, pages 1–13, 2018.

[106] Mirja Peltola. Role of editing of rr intervals in the analysis of heart rate variability. *Frontiers in physiology*, 3:148, 2012.

[107] Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods*, 17(3):261–272, 2020.

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

[109] Chad Harms and Frank Biocca. Internal consistency and reliability of the networked minds measure of social presence. 2004.

[110] João Maroco, Ana Lúcia Maroco, Juliana Alvares Duarte Bonini Campos, and Jennifer A Fredricks. University student's engagement: development of the university student engagement inventory (usei). *Psicologia: Reflexão e Crítica*, 29, 2016.

[111] Yan He, Dingding Zheng, George. Chernyshov, Ragnar. Thomsen, Jiawen. Han, Danny Hynds, Yun Suen. Pai, Kai. Kunze, and Kouta. Minamizawa. Frisson waves: Sharing frisson to create collective empathetic experiences for music performances. In *2021 IEEE World Haptics Conference (WHC)*, pages 591–591, 2021. `doi:10.1109/WHC49131.2021.9517258`.

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

[113] Sarah Fdili Alaoui, Kristin Carlson, and Thecla Schiphorst. Choreography as mediated through compositional tools for movement: Constructing a historical perspective. In *Proceedings of the 2014 International Workshop on Movement and Computing*, pages 1–6, 2014.

[114] A John Camm, Marek Malik, J Thomas Bigger, Günter Breithardt, Sergio Cerutti, Richard J Cohen, Philippe Coumel, Ernest L Fallen, Harold L Kennedy, RE Kleiger, et al. Heart rate variability: standards of measurement, physiological interpretation and clinical use. task force of the european society of cardiology and the north american society of pacing and electrophysiology. 1996.

[115] Jerzy Sacha and Władysław Pluta. Alterations of an average heart rate change heart rate variability due to mathematical reasons. *International journal of cardiology*, 128(3):444–447, 2008.

[116] Jerzy Sacha. Why should one normalize heart rate variability with respect to average heart rate. *Frontiers in physiology*, 4:306, 2013.

[117] Solange Akselrod, David Gordon, Jeffrey B Madwed, Nancy C Snidman, Danied C Shannon, and Richard J Cohen. Hemodynamic regulation: in-

vestigation by spectral analysis. *American Journal of Physiology-Heart and Circulatory Physiology*, 249(4):H867–H875, 1985.

[118] JPA Delaney and DA Brodie. Effects of short-term psychological stress on the time and frequency domains of heart-rate variability. *Perceptual and motor skills*, 91(2):515–524, 2000.

[119] David Nunan, Gavin RH Sandercock, and David A Brodie. A quantitative systematic review of normal values for short-term heart rate variability in healthy adults. *Pacing and clinical electrophysiology*, 33(11):1407–1417, 2010.

[120] Tom Kuusela. Methodological aspects of heart rate variability analysis. *Heart rate variability (HRV) signal analysis: Clinical applications*, pages 10–42, 2013.

[121] Stephen W Porges. Vagal tone: An autonomic mediator of affect. 1991.

[122] Sybil Huskey, Celine Latulipe, Melissa Word, and Danielle Lottridge. Post-production focus groups in dance: A case study and protocol. *Journal of Dance Education*, 18(2):47–54, 2018.

[123] Takuya Iwamoto and Soh Masuko. Lovable couch: Mitigating distrustful feelings for couples by visualizing excitation. In *Proceedings of the 6th Augmented Human International Conference*, pages 157–158, 2015.

[124] Henrique Sequeira, Pascal Hot, Laetitia Silvert, and Sylvain Delplanque. Electrical autonomic correlates of emotion. *International journal of psychophysiology*, 71(1):50–56, 2009.

[125] 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.

[126] Sarah Fdili Alaoui. Making an interactive dance piece: tensions in integrating technology in art. In *Proceedings of the 2019 on Designing Interactive Systems Conference*, pages 1195–1208, 2019.

[127] Celine Latulipe, David Wilson, Berto Gonzalez, Adam Harris, Erin Carroll, Sybil Huskey, Melissa Word, Robert Beasley, and Nathan Nifong. Soundpainter. In *Proceedings of the 8th ACM conference on Creativity and cognition*, pages 439–440, 2011.

[128] Graeme Sullivan. *Art practice as research: Inquiry in visual arts.* Sage, 2010.

[129] Hazel Smith. *Practice-led research, research-led practice in the creative arts.* Edinburgh University Press, 2009.

[130] Félix Schoeller, AJH Haar, Abhinandan Jain, and Pattie Maes. Enhancing human emotions with interoceptive technologies. *Physics of life reviews*, 31:310–319, 2019.

[131] Scott Bannister. Distinct varieties of aesthetic chills in response to multimedia. *PloS one*, 14(11):e0224974, 2019.

[132] Rémi de Fleurian and Marcus Pearce. Chills in music: An integrative review. *PsyArXiv*, 1:1–50, 2020.

[133] Oliver Grewe, Reinhard Kopiez, and Eckart Altenmüüller. The chill parameter: Goose bumps and shivers as promising measures in emotion research. *Music Perception*, 27(1):61–74, 2009.

[134] Mathias Benedek and Christian Kaernbach. Physiological correlates and emotional specificity of human piloerection. *Biological psychology*, 86(3):320–329, 2011.

[135] Kejun Zhang, Hui Zhang, Simeng Li, Changyuan Yang, and Lingyun Sun. The pmemo dataset for music emotion recognition. In *Proceedings of the 2018 acm on international conference on multimedia retrieval*, pages 135–142, 2018.

[136] Ella A Cooper, John Garlick, Eric Featherstone, Valerie Voon, Tania Singer, Hugo D Critchley, and Neil A Harrison. You turn me cold: evidence for temperature contagion. *PloS one*, 9(12):e116126, 2014.

[137] Kristin Neidlinger, Khiet P Truong, Caty Telfair, Loe Feijs, Edwin Dertien, and Vanessa Evers. Awelectric: that gave me goosebumps, did you feel it too? In *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction*, pages 315–324, 2017.

[138] Wolfgang Tschacher, Steven Greenwood, Hauke Egermann, Melanie Wald-Fuhrmann, Anna Czepiel, Martin Tröndle, and Deborah Meier. Physiological synchrony in audiences of live concerts. *Psychology of Aesthetics, Creativity, and the Arts*, 2021.

[139] Kelsey E Onderdijk, Dana Swarbrick, Bavo Van Kerrebroeck, Maximillian Mantei, Jonna K Vuoskoski, Pieter-Jan Maes, and Marc Leman. Livestream experiments: the role of covid-19, agency, presence, and social context in facilitating social connectedness. *Frontiers in psychology*, 12:647929, 2021.

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

[141] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

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

[143] Gary G Berntson, Karen S Quigley, Greg J Norman, and David L Lozano. Cardiovascular psychophysiology. 2017.

[144] Haoshi Zhang, Mingxing Zhu, Yue Zheng, and Guanglin Li. Toward capturing momentary changes of heart rate variability by a dynamic analysis method. *PLoS One*, 10(7):e0133148, 2015.

[145] Lauri Ahonen, Benjamin Cowley, Jari Torniainen, Antti Ukkonen, Arto Vihavainen, and Kai Puolamäki. Cognitive collaboration found in cardiac

physiology: Study in classroom environment. *PloS one*, 11(7):e0159178, 2016.

[146] Jorina von Zimmermann, Staci Vicary, Matthias Sperling, Guido Orgs, and Daniel C Richardson. The choreography of group affiliation. *Topics in Cognitive Science*, 10(1):80–94, 2018.

[147] Claudio Martella, Ekin Gedik, Laura Cabrera-Quiros, Gwenn Englebienne, and Hayley Hung. How was it? exploiting smartphone sensing to measure implicit audience responses to live performances. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 201–210, 2015.

[148] P Van Gent, H Farah, N Nes, and B van Arem. Heart rate analysis for human factors: Development and validation of an open source toolkit for noisy naturalistic heart rate data. In *Proceedings of the 6th HUMANIST Conference*, pages 173–178, 2018.

Appendices

A. Glossary

The definitions are summarized referring to either Wikipedia or related papers [52, 61] and ordered alphabetically.

Autonomic Nervous System (ANS): The autonomic nervous system (ANS) is a division of the peripheral nervous system that supplies smooth muscle and glands, and thus influences the function of internal organs. ANS is a control system that acts largely unconsciously and regulates bodily functions.

Band-pass filter: A band-pass filter allows through components in a specified band of frequencies, called its passband but blocks components with frequencies above or below this band.

Biofeedback: Biofeedback is the process of gaining greater awareness of many physiological functions of one's own body. In this dissertation, it is defined as using physiological data to create interactive experiences.

Choreography: Choreography is the art or practice of designing sequences of movements of physical bodies (or their depictions) in which motion or form or both are specified.

Dynamic time warping: In time series analysis, dynamic time warping (DTW) is an algorithm for measuring similarity between two temporal sequences.

Electrodermal activity (EDA): The umbrella term used for defining autonomic changes in the electrical properties of the skin. The most widely studied property is the skin conductance, which can be quantified by applying an electrical potential between two points of skin contact and measuring the resulting current flow between them.

Frisson: Frisson is a psycho-physiological phenomenon commonly described as having goosebumps, or feeling shivers down one's spine, that can be triggered from external stimuli such as music or intense emotions.

Fortissimo: Fortissimo is borrowed from an Italian word that means very loud. It represents dynamic or volume level in western music indicating the piece is played very loud. **heart rate variability (HRV):** Heart rate variability (HRV) is the fluctuation in the time intervals between adjacent heartbeats.
HF power: The power of the high-frequency band (0.15–0.4 Hz)
High-pass filter: A high-pass filter is an electronic filter that passes signals with a frequency higher than a certain cutoff frequency and attenuates signals with frequencies lower than the cutoff frequency.
IDE: An integrated development environment (IDE) is a software application that provides comprehensive facilities to computer programmers for software development.
Interbeat Interval (IBI): Time interval between successive heartbeats.
LF power: The power of the low-frequency band (0.04–0.15 Hz).
LF/HF: The ratio of LF to HF power.
Low-pass filter: A low-pass filter is a filter that passes signals with a frequency lower than a selected cutoff frequency and attenuates signals with frequencies higher than the cutoff frequency.
NN intervals: Interbeat intervals from which artifacts have been removed.
OSC: Open Sound Control (OSC) is a protocol for networking sound synthesizers, computers, and other multimedia devices for purposes such as musical performance or show control.
Parasympathetic Nervous System (PSNS): The parasympathetic nervous system (PSNS) is one of the ANS divisions and is responsible for stimulation of "rest-and-digest" activities that occur when the body is at rest.
pNN50: Percentage of successive RR intervals that differ by more than 50 ms. One HRV feature that is strongly correlated with the activation of PSNS.
RMSSD: Root mean square of successive RR interval differences.
RR intervals : Interbeat intervals between all successive heartbeats.
SDNN : Standard deviation of NN intervals.
support-vector machines (SVM): In machine learning, support-vector machines (SVMs) are supervised learning models with associated learning algorithms that analyze data for classification and regression analysis.
sympathetic nervous system (SNS): The sympathetic nervous system (SNS)

is one of the ANS divisions and is to stimulate the body's fight or flight response.
TCP/IP: Transmission Control Protocol (TCP) is one of the main protocols of the Internet protocol suite. It originated in the initial network implementation in which it complemented the Internet Protocol (IP). Therefore, the entire suite is commonly referred to as TCP/IP
UDP: User Datagram Protocol (UDP) is one of the core members of the Internet protocol suite. With UDP, computer applications can send messages, in this case referred to as datagrams, to other hosts on an Internet Protocol (IP) network.

B. Large Scale Dataset

There are two large scale dataset collected in-the-wild analyzed in this thesis. We publicize the two dataset on the open-source website for those who may conduct similar research.

B.1 Dataset from Boiling Mind Project

The dataset consists of audience multi-modal signals (EDA, BVP, wrist acceleration, and angular velocity) over three performances.

In performance 1, we have 34 recordings (male =17; female =17). In performance 2, we have 31 recordings (male =13; female =18). In performance 3, we have 33 recordings (male =19; female =14).

The dataset is stored at:https://osf.io/sypz4/.

B.2 Dataset from Frisson Waves Project

We have 33 recordings (EDA, BVP, and frisson labels) in total (male = 16; female = 17) from the lab study to develop frission detection model. In this thesis, we removed incomplete and noisy data leaving 19 participants‘ data. We have 48 recordings collected (BVP and EDA data) from in-the-wild conert (male = 19; female = 28, prefer not to say=1).

The dataset together with the questionnaires is stored at: https://osf.io/rzpn3/.

C. Example Codes for Physiological Data Analysis

This section provides example codes written in python for feature extraction in offline analysis and real-time analysis. The codes are developed using multiple python packages, such as *scipy* [107], *heartpy* [148], and *Neurokit2* [100]. Using those established packages could help extract HRV features and EDA features in a fast and feasible way.

C.1 Example codes for offline analysis

The codes in this section use BVP data as an example to explain the feature extraction process.

```
import numpy as np
import pandas as pd
from scipy.signal import butter, filtfilt
import heartpy as hp
import neurokit2 as nk

# functions
def set_timepass(data, timename, start):
    data['timepass'] = round(data[timename] - start)
    data['timepassMin'] = round(data['timepass']/60, ndigits= 0)
    return data

def lowpass(data, signaltype, freq, cutoff, order):
    w = cutoff /(freq/2)
    b,a = butter(order, w,'low')
    filtered_value =  filtfilt(b,a,data[signaltype])
    return filtered_value

def bvp_adjust(raw_list, samplerate):
    filtered_list = hp.remove_baseline_wander(raw_list, samplerate)
    scaled_list = hp.scale_data(filtered_list, lower =0,
                    upper = 2000)
    return scaled_list
```

```
def get_hrv_features(data, timetype, window, increment, adjust):
    HRV_features = ['HRV_MeanNN', 'HRV_SDNN', 'HRV_RMSSD',
                    'HRV_SDSD', 'HRV_pNN50', 'HRV_pNN20', 'HRV_LFHF']
    hrv_list = []
    timepassSec= []

    i = data[timetype].min()
    while True:
        signalok = (data[timetype] >= i)&
                    (data[timetype] < i + window)

        if signalok.sum() >= 200:
            raw_list = data['bvp_filtered'][signalok]

            if adjust == True:
                bvp_list = preprocess.bvp_adjust(raw_list,100)
            else:
                bvp_list = raw_list

            info =  nk.ppg_findpeaks(hp.enhance_peaks(bvp_list),
                    sampling_rate=100, method = 'elgendi')
            peaks = info['PPG_Peaks']
            peaks = peaks[np.logical_not(np.isnan(peaks))]
            hrv = nk.hrv(peaks, sampling_rate =100,
                        show = False)[HRV_features]
            i = i + increment
            hrv_list.append(hrv)
            timepassSec.append(i+increment)
        else:
            i = i + increment
        if i > data[timetype].max() - window:
            break
    return hrv_list,timepassSec

def get_hrv_dataframe(data, timetype, window, increment, adjust):

    hrv = get_hrv_features(data, timetype, window,
                    increment, adjust)[0]
    timepassSec = get_hrv_features(data, timetype,
```

```
                    window, increment, adjust)[1]

    concatelist = []
    for i in range(len(hrv)):
        concatelist.append(hrv[i])

    hrv_dataframe =pd.concat(concatelist, axis=0).reset_index()
    hrv_dataframe.interpolate(inplace=True)
    hrv_dataframe[‘timepassSec’] = timepassSec
    hrv_dataframe = hrv_dataframe.groupby(
                    'timepassSec'
                    ).mean().reset_index()

    return hrv_dataframe

# execute the code
# df_bvp --> dataset saved in the format of dictionary
# earliestTime --> the start time of the group event

for key in sublist:
    df_bvp[key] = set_timepass(df_bvp[key], ‘localTime’,
                               earliestTime)
    df_bvp[key][‘bvp_filtered’] = preprocess.lowpass(df_bvp[key],
                                   ‘bvp’,200,3,2)

hrv_dataframe= dict()
for key in sublist:
    print(key)
    hrv_dataframe[key] = feature_extract.get_hrv_dataframe(
                         df_bvp[key],‘timepassSec’,
                         240,120,True)
    hrv_dataframe[key][‘subject_number’] = key

scaler = MinMaxScaler()
for key in sublist:
    for feature in hrv_features:
        data = hrv_dataframe[key][feature].to_numpy()
        data = data.reshape((data.shape[0],1))
        hrv_dataframe[key][‘normalized_’+ feature] =
                               scaler.fit_transform(data)
```

```
#hrv_dataframe > features saved in the format of dictionary
```

C.2 Example codes for real-time analysis

The codes in this section use EDA data as an example to explain the feature extraction process. As an example for real-time analysis, codes to stream data and implement the trained model are also included.

```
import sys
import time
import pandas as pd
import numpy as np
import neurokit2 as nk
from scipy.signal import butter, filtfilt
from sklearn.preprocessing import MinMaxScaler, StandardScaler
import pickle
import random
pd.options.mode.chained_assignment = None
from random import seed
from random import random

#functions
def lowpass(data, freq, cutoff, order):
    w = cutoff/(freq/2)
    b, a = butter(order, w, 'low')
    data_filtered = filtfilt(b, a, data)
    return data_filtered

def get_eda_features(data):
    data['eda_filtered_good'] = lowpass(df_eda['eda'],
                                        4.545, 0.01, 2)
    data['eda_filtered_good'] = data['eda_filtered_good'].fillna(0)
    signals, info = nk.eda_process(data['eda'],
                                        sampling_rate= 4.545)
    eda_dataframe = signals[['EDA_Tonic', 'EDA_Phasic']]
    eda_dataframe = eda_dataframe.fillna(0)

    eda_dataframe['ok'] = 1
    scaler = MinMaxScaler(feature_range = (0,1))
```

```
    for feature in ['EDA_Tonic', 'EDA_Phasic']:
        data = eda_dataframe[feature].to_numpy()
        data = data.reshape((data.shape[0],1))
        eda_dataframe['normalized_'+ feature] =
                        scaler.fit_transform(data)

    return(eda_dataframe.tail(1))


# execute code

df_eda = pd.DataFrame()
dataBuffer_eda = []
seed()

with open('ensemble_Model.pkl', 'rb') as file:
    ensemble_Model = pickle.load(file)
    file.close()

with open('scaler_model.pkl', 'rb') as file:
    scaler_Model = pickle.load(file)
    file.close()

sys.stdout.write(''Python Starting\n")
sys.stdout.flush()


for line in sys.stdin:
    # Remove trailing newline characters using strip()
    if 'exit' == line.strip():
        sys.stdout.write('Found exit. Terminating the program\n')
        sys.stdout.flush()
        exit(0)
    else:
        if(line[0] == 'e'):
            dataBuffer_eda.append(float(line[1:]))

        else:
            dataBuffer_bvp.append(float(line[1:]))
```

```
            if(len(dataBuffer_eda) == 250):
                startTime = time.perf_counter()
                df_eda = df_eda.append(pd.DataFrame(dataBuffer_eda,
                        columns=['eda']), ignore_index=True)
                dataBuffer_eda = []

                if(len(df_eda) > 300):
                    df_eda = df_eda.iloc[len(df_eda)-300:]

                feature_dataframe = get_eda_features(df_eda)
                feature_dataframe = feature_dataframe.fillna(0);

                # making predictions
                X = feature_dataframe[['normalized_EDA_Tonic',
                    'normalized_EDA_Phasic']]
                X_scaled =  scaler_Model.transform(X)
                y = ensemble_Model.predict(X_scaled)

                sys.stdout.write(str(y[0]))
                sys.stdout.write(",”+str(time.perf_counter()
                                - startTime))
                sys.stdout.write("\n”)
                sys.stdout.flush()
sys.stdout.write("Python Stopping\n”)
sys.stdout.flush()
```