UNIVERSITY OF CALIFORNIA
SANTA CRUZ

ALL THE FEELS: A TWITCH OVERLAY THAT DISPLAYS STREAMERS' BIOMETRICS TO SPECTATORS

A thesis submitted in partial satisfaction of the
requirements for the degree of

MASTER OF SCIENCE

in

COMPUTATIONAL MEDIA

by

Raquel Robinson

March 2018

The Thesis of Raquel Robinson
is approved:

Professor Katherine Isbister, Chair

Professor Noah Wardrip-Fruin

Professor Regan Mandryk

Tyrus Miller
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

ALL THE FEELS: A TWITCH OVERLAY THAT DISPLAYS STREAMERS' BIOMETRICS TO SPECTATORS

by

Raquel Robinson

All the Feels is a tool I created that provides an overlay of biometric and webcam-derived data onto the interface of the popular video game streaming service, Twitch. This overlay provides visualization of heart rate, skin conductance, and facial emotion recognition. It is intended for two purposes: one, to enhance the spectator experience and improve streamer-spectator connection and two, to evaluate players' emotional response to games. I present results of a few preliminary studies (phase 1 and phase 2) investigating reactions to the introduction of this tool in various environments through the Twitch streaming platform. The first two studies are accounts of results via live-testing with streamers and spectators, and the third is with the tool used in combination with other methods to evaluate player experience in games.

This work opens up opportunities to improve the streaming experience and spectator engagement through the introduction of an interface that includes biometric signals, and also has broader applications as a tool for understanding player experience in general. The thesis describes this tool both as a way of evaluating players' emotional response to games, and also as a way for spectators to see players' emotional response

to games. I discuss the implementation in terms of these two different contexts, while providing ideas for future use of this tool in games user research.

I dedicate this work to my one true love, gameing [1].

Dear gaming,

From the moment I yelled my first slur in Halo And shooting imaginary Game-winning head-shots In the esports arena I knew one thing was real: I fell in love with you. A love so deep I gave you my all – From my mind & avatar To my spirit & persona. As a six-year-old girl gamer Deeply in love with you I never saw the end of the level. I only saw myself Beating Level one. And so I gamed. I gamed up and down every map After every loot box for you. You asked for my APM I gave you my wrists Because it came with so much more. I played through the sweat and doritos dust Not because challenge called me But because YOU called me. I did everything for YOU Because that's what you do When someone makes you feel as Alive as you've made me feel. You gave a six-year-old girl gamer her Gaming dream And I'll always love you for it. But I can't love you obsessively for much longer. This competitive season is all I have left to give. My heart can take the pounding My mind can handle the grind But my body knows it's time to say goodbye. And that's OK. I'm ready to let you go. I want you to know now So we both can savor every moment we have left together. The good and the bad. We have given each other All that we have. And we both know, no matter what I do next I'll always be that kid With the fedora Garbage can in the corner :05 seconds on the clock Controller in my hands. Ready ... 3 ... 2 ... 1 ... GO

Love you always,

Raquel

[1] Based off Kobe Bryant's poem: Dear Basketball

Acknowledgments

The text of this thesis includes reprints of the following previously published material:

Raquel Robinson, Katherine Isbister, and Zachary Rubin. 2016. All the Feels: Introducing Biometric Data to Online Gameplay Streams. In CHI PLAY '16 Extended Abstracts, 261-267.

Raquel Robinson, Zachary Rubin, Elena Márquez Segura, and Katherine Isbister. 2017. All the Feels: Designing a Tool That Reveals Streamers' Biometrics to Spectators. In Proceedings of the 12th International Conference on the Foundations of Digital Games (FDG '17), 36:1–36:6.

Raquel Robinson, John Murray, and Katherine Isbister. 2018. "You're Giving Me Mixed Signals!": A Comparative Analysis of Methods that Capture Players' Emotional Response to Games. In Forthcoming CHI '18 Extended Abstracts.

The co-author Katherine Isbister, listed in these publications, directed and supervised the research which forms the basis for the thesis. The co-author Zachary Rubin helped with programming, study design and execution, and paper editing. The co-author Elena Márquez-Segura helped with paper framing, positioning within the field of HCI, as well as general paper guidance and mentoring. The co-author John Murray contributed some of the dataset annotation strategy (using Premiere Pro), preparation (video editing to get things in sync), and some general collaboration on consolidating the protocol for conducting the study.

Chapter 1

Introduction

Capturing users' emotional experience has always been a challenge for human-computer interaction researchers. It is especially challenging in the domain of games, because games are aimed at creating a broad range of emotional experiences [11] that existing tools for evaluating other digital applications were not designed to capture, and so might miss or distort. Over the years, researchers in the player experience domain have adapted and developed various techniques to help evaluate these emotional experiences. A mixture of self-report questionnaires and biometrics are typically used for this purpose [21]. Additionally, biometric sensor technology in recent years has become a lot more advanced, widespread, and affordable, making it more feasible to use in studies aimed at evaluating player experience. This work presents a tool called All the Feels (ATF) which I developed as an unobtrusive way to capture players' emotional response to games. It displays heart rate and GSR (galvanic skin response) information, taken from the player in real-time, through the use of a commercially available wearable device. The tool also provides emotion labeling information input from existing auto-detection software that uses the player's webcam. I use this tool: a. As a way to increase engagement among Twitch live stream spectators and b. During a comparative analysis of methods for capturing player experience. I explore this tool with respect to two sides of a coin: player experience and spectator experience. I made it both as a way of evaluating players' emotional response to games, as well as a way for spectators to see players' emotional response to games. In this thesis, I discuss the implementation in terms of these two different contexts, while providing ideas for future use of this tool in games user research.

a. In this context, I explore the potential value of introducing additional insight to spectators regarding the player's emotional state, by including a Heads-up display (HUD) that provides biometric and facial detection information in real time. All The Feels is conceived of as an additional playful set of information that could increase spectators' engagement, enjoyment, and emotional connection with the player-streamer. The intent is to use an interactional model of emotion rather than an informational one — focusing on creating engaging and flexible dialog and interaction among streamers and spectators [4] rather than seeking to adapt and respond to emotion through changes in game state. This section details the design process exploring this idea in two phases: first by creating an early prototype of the tool and conducting a small-scale user study (phase 1), and then creating a more robust version tested with a real streamer in a two-hour long stream (phase 2). b. In this context, I use All the Feels along with several other methods to gather insights into how different techniques compare in registering players' emotional responses to a game. Many techniques are used to understand and evaluate emotions about games. Classic techniques used to evaluate player experience include think-aloud, post-game questionnaires, and retrospective interviews with players about the game experience. Newer and less-explored methods have been developed to capture a broader range of emotions that these traditional techniques can't capture on their own. The Sensual Evaluation Instrument (SEI) is among these; a set of eight objects varying in size and shape for the player to hold, which gives them a way to express emotions without using words [12]. However, no comparison of these techniques in the context of gaming has been conducted to date. Which of these methods are useful in which contexts? Are there differences among these methods that make them better or worse for understanding player feelings associated with certain moments in games? This section explains my exploration of these questions by conducting a comparative analysis of several player experience evaluation techniques, including All the Feels, post-play questionnaires, and other techniques, in a lab-based study. Results support using a mixed-methods approach when evaluating players' emotional response [30].

Chapter 2

Live-Streaming

2.1 Background

Since the advent of televised digital gaming in the 1990s, sharing gameplay video footage has been a fundamental part of gaming culture and activity. In the last decade, video sharing services such as YouTube have become more prominent. Additionally, esports (playing computer games in tournaments with spectators) has become an emerging phenomenon, driving interest in sharing gameplay both live and recorded. The intense popularity of the streaming phenomenon in recent years is largely due to the creation of Twitch. Twitch was founded in 2011 and since has become the leading service for gameplay streaming. The popularity of gameplay streaming continues to rise, with 100+ million unique viewers per month and 1.7+ million broadcasters per month [1]. The videogame livestreaming market continues to grow with sites such as YouTube creating communities of their own.

A related phenomenon is Let's Play videos. These game walkthrough videos on YouTube can be traced back to 2005 on the internet forum 'Something Awful'. Users on this forum would comment and participate in a video game, telling the OP — original poster — where to go next and their thoughts and commentary on what was happening. The OP would keep the community updated about the game with screenshots and commentary as well [15]. Many players now use streaming services like Twitch to broadcast live Let's Plays. In these play streams, the voiceover or webcam video adds both information and emotional nuance to the spectator's experience. These tools have opened up a new channel of communication between players and spectators.

Spectators using Twitch do not simply passively view a player's stream, they also can interact with the streamer during play. The Twitch platform allows spectators to chat using both typed text and emoji, engaging in lively side conversation in real time. As Stenros [35] points out, sociability that occurs alongside play like this can have an important impact on the experience of games and is an important design consideration in its own right. Over the last decade, biometric sensors have of course been explored in the context of game design and research. Sensors have primarily been used as novel (interaction) inputs to games e.g [26], and as to tools to evaluate games and their effect in players [22, 23]. The present work takes a different direction, examining how such sensors can enhance the spectator experience. In this regard, research and design in the domain of interactive performances can be illuminating, for example Reeves et al.'s investigation of spectator experience of interactive installations [28], and Wang et al.'s work assessing audience engagement with a theater performance [39]. Tennent et al.

[37] engaged in a similar project during a prototype TV show, displaying physiological data from actors. The goals of these authors, to allow the audience to engage in "sense-making practices when watching actors" and "potentially enhance the vicarious nature of viewing another's experience", resonate with our goal of providing a rich spectator experience. The same research team has also implemented a tool for capturing video, audio, heart-rate, and acceleration data of those riding a roller coaster, streaming this live to spectators, thus transforming riders into performers [33].

In the realm of game design, biometrics to enrich the spectator experience were used during Dreamhack, a biannual European esports event where key players around the world gather to play competitive games that are broadcasted on a large scale. In 2011, Dreamhack equipped players with heart rate monitors while streaming games, which was commented upon and celebrated by the audience: "I think it's a great addition and I'm really getting more into the player's minds by knowing how stressed they are in the particular moment." [27] Unfortunately, subsequent Dreamhack events discontinued the display of this biometric feature, which some fans attributed to distraction for the players [27]. This intervention shows the potential of displaying the players' biometrics for the sake of a rich spectator experience. Yet, this is an under-explored topic that needs to be further examined.

2.1.1 Sociability Elements in Twitch

In recent years, there has been a rise in the popularity of gameplay streaming platforms, due in part to decreases in both the costs of streaming tools and the knowl-

Figure 2.1: Wgrates uses a dog avatar to engage spectators on his Twitch stream.

edge required to operate them. The most popular platform, Twitch, allows streamers a variety of customization opportunities. Using live stream configuration tools such as OBS, FFsplit and Xsplit, streamers can add music, advertisements, and other data to the display of their gameplay.

Customizability combined with robust chat features makes Twitch a fertile environment for experiments in sociability. For example, Twitch Plays Pokémon turned the chat room feature into a game controller, and allowed users around the world to collaborate to advance in the game [2]. Through clever use of green screening and facial recognition, Twitch streamer wgrates transforms himself into a dog sitting at a desk, floating in a corner of the screen. The dog's mouth and face moves in accordance with wgrates's speech and facial expressions (see Figure 2.1).

2.2 Phase 1: Early Prototype of All the Feels

Our goal was to gather relevant and engaging biometric data in unobtrusive ways so as not to affect gameplay. To better achieve this goal we organized a requirements gathering session, conducted an analysis of existing commercial devices, and did some informal user testing of a subset of devices.

2.2.1 Requirements Gathering Session

We held a focus group with eight graduate students recruited via a human-computer interaction course offered at the Baskin School of Engineering at the University of California, Santa Cruz to help us gather requirements for the tool. We provided participants with the Empatica and Muse biometric wearables as examples of potential devices, and set up a game stream using the video game Thomas Was Alone. Participants tried out the equipment while streaming, then provided feedback about their experience. Based on the feedback we determined several requirements: 1) It was crucial that our tool integrate well with existing streaming software. 2) Participants were more interested in heart rate and GSR than in brainwave (EEG) data. 3) Participants wanted wearables that were easy to get hold of commercially, and a tool that was compatible with Microsoft Windows. 4) Latency was a concern that needed to be prioritized.

2.2.2 Device Analysis

We investigated commercially available devices that provide different sets of biometric data including heart rate variability (HRV), galvanic skin response (GSR), and electroencephalogram (EEG) (see Table 2.1).

Our primary design goal was to maximize the number of unique biometrics while minimizing streamer discomfort. Subsequently, we elected to maximize interop-

Name	Video	Location of Device	GSR	HRV
Empatica	No	Worn, Wrist	Yes	No
Affdex	Yes	Webcam	No	No
Microsoft Band	No	Worn, Wrist	Yes	Yes
Apple Watch	No	Worn, Wrist	No	Yes
Muse	No	Worn, Head	No	No

Table 2.1: Comparison of devices that provide biometric data

erability between devices and remove devices that led streamers to express discomfort. Participants found the Muse, a headband which tracks EEG data, to be the most unusable device. Participants with glasses were incapable of using the device while wearing glasses. EEG data was found to be extremely unreliable and noisy — consistent with other studies [20]. The Microsoft Band is both compatible with Windows and has GSR and HR biometrics. After this hardware investigation, we came to the conclusion that the Microsoft Band and Affdex best fit the requirements. Therefore the first prototype overlay consisted of the Affdex data in a corner of the webcam display, and the Microsoft band heart rate and GSR data below it (see Figure 2.2). (Affdex is facial expression analysis software that detects and labels emotions in real-time using the webcam stream.)

2.2.3 User Test of Prototype

The purpose of this initial study was to have spectators and streamers test out the prototype and get their initial impressions and reactions of its potential, as well as interact and network with the streamer/spectator community. Another motivation was to figure out common streaming setups in hopes of integrating the tool into the

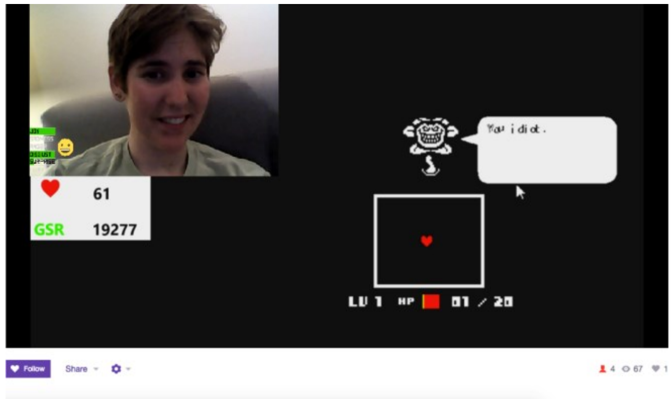

Figure 2.2: Heads up display on Twitch of the High Fidelity Prototype of All the Feels. Affdex data is shown within the player's video window using emotion labels and icons (top left). Microsoft Band (GSR and HR) data is displayed below the video window. The game takes up most of the streaming display.

streamers' existing setups in the future for studies, rather than using one specific laptop for testing.

2.2.4 Procedure

Participants included two females and three males aged 20-30 years old. All male participants reported that they watch streams regularly; one reported that he watches and streams for professional purposes — that is to say he watches other players streams for competitive purposes. One female participant watches let's play videos, the other does not watch streams.

The study sessions took place at two different times and locations. The first location was with one spectator and one streamer, and the second was with two spectators and one streamer. In all cases, spectators sat in a different room from the streamer. Computational resources taken up by the streaming software limited us to light 2D/3D

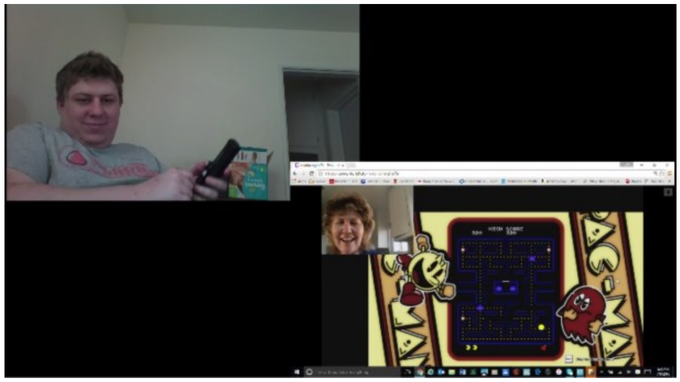

Figure 2.3: Spectator reactions to streaming overlay without the addition of emotional data. Spectator is in the upper left and stream is in the lower right.

games. Games played included Bit Trip Runner and Pac-man. Each subject took turns being a player streaming themselves playing and a spectator. Play sessions lasted five minutes each. First, we had the player stream normally without using the tool. The stream overlay included a webcam display in the upper left corner, and the game on the rest of the screen.

Next, we activated the tool and had the streaming player continue. During the entire test, we recorded the spectator's view (Figure 2.3), and we also recorded spectators using a webcam trained on them. We combined these videos with the stream videos to conduct our analysis (see Figures 2.4 and 2.5).

After each participant got the chance to stream and spectate, they answered questions in a paper questionnaire format, from both the spectator and player points of view. Questions about playing were aimed at finding out whether wearing the Microsoft Band and knowing their emotions were being exposed changed gameplay. Questions about spectating included whether or not the biometric information displayed was

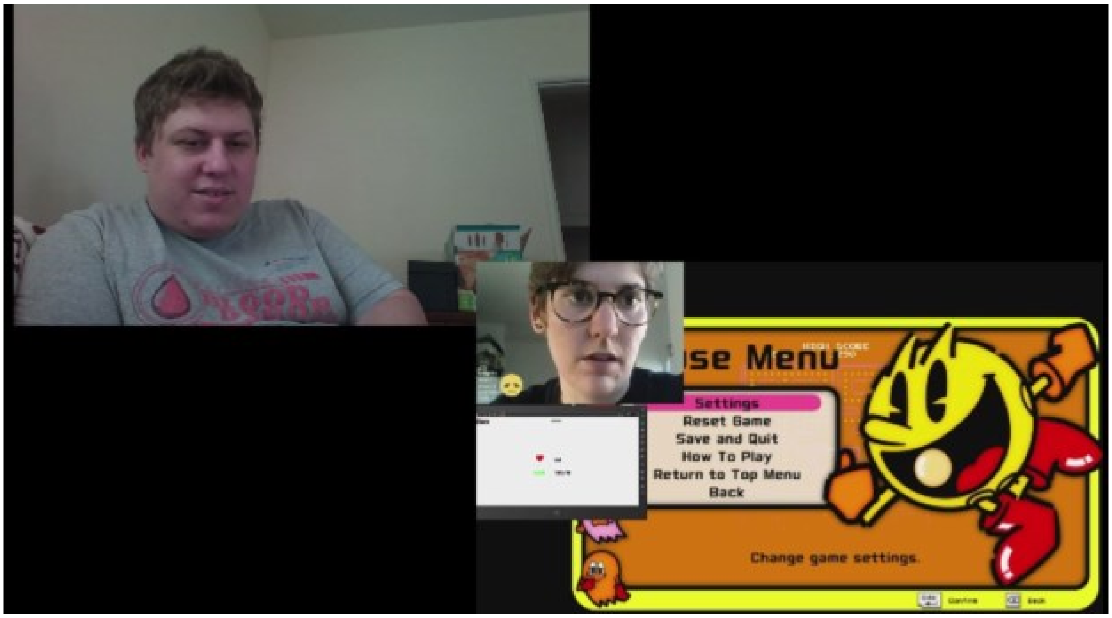

Figure 2.4: Spectator reaction to stream with emotional data. Spectator is in the upper left and stream is on the lower right.

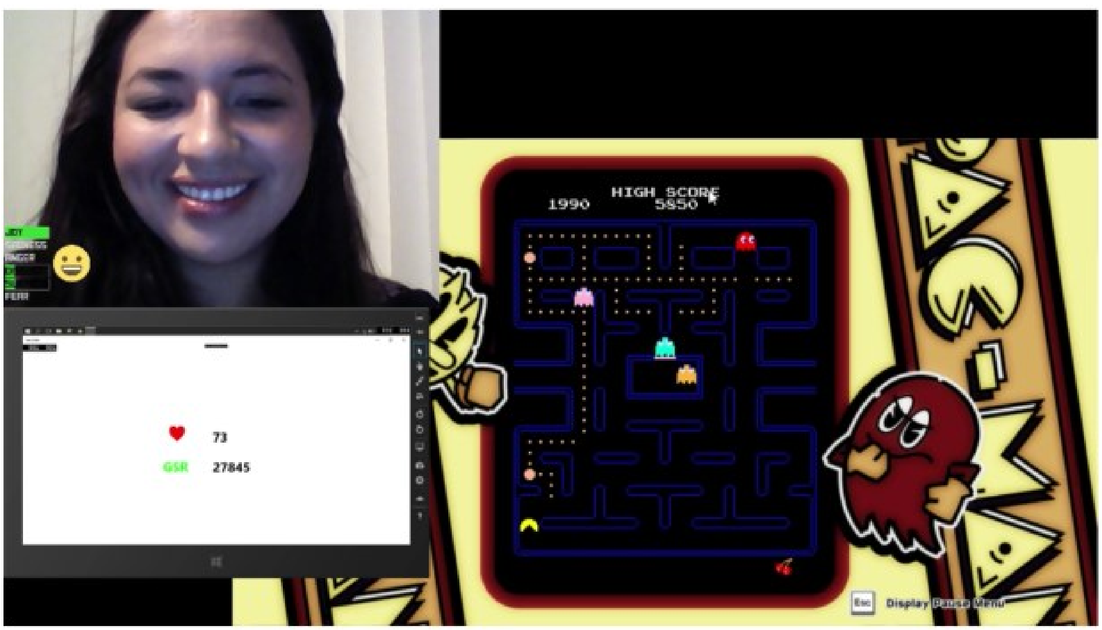

Figure 2.5: Spectator perspective of the stream via Twitch.

engaging, overwhelming, or interesting, and if they would make any changes to the interface.

From the streamers' perspective, wearing the Microsoft Band did not negatively impact the experience. It did not impede their range of motion in gameplay. The need to be more expressive because they knew their emotions were being tracked was stated by three of five participants. Three players also noted that this tool would be useful for stress regulation while playing games. To know how they react to certain situations would come in handy for self-reflection and the improved ability to regulate emotional responses to events in gameplay. One participant said "Maybe it is a good indicator of when to stop playing when feeling frustrated."

It's important to note that the Microsoft Band presented some challenges in the study. Only the small size of the Band was available for use during the study (it comes in medium and large as well), and two of the five participants were too large to wear the device comfortably. Heart rate information was sometimes inaccurate, and the device had difficulty tracking heart rate above 100 beats per minute.

A key finding was that female participants had more interest in the tool than male participants. Both female participants reported more empathy toward the player than male participants did. One female participant was quoted as saying, "It improved my experience because I felt more connected to the player emotionally." At the time of writing females make up only 20% of the online game streaming user-base. Use of this tool like our prototype could perhaps contribute to increasing female presence on game streaming platforms.

All male participants reported that while they did not experience an increased empathy toward the streamer, they did have interest in the biometric data. Two of the three male participants said that the tool would be best used for casual streaming, and would not be of interest in competitive play streams or esports. One male participant said this would be interesting for all game streams, both casual and competitive. Another male participant was quoted as saying, "I feel like this would get gimmicky if used in a professional setting."

The laptop used for the study could not stream and play graphically intense games at the same time. As a result, participants were unable to play a game of their choosing. Spectators did not feel that Pac-man was a good choice for testing the tool, because it was not terribly engaging and emotional. Horror, fighting, and hide-and-go seek type games were suggested as interesting genres of games to use with the tool instead. For future studies, we planned to use games that evoke greater emotional responses in streamers.

One of the participants pointed out that some who watch streams do so while engaging in other activities — keeping the stream on in the background. This person suggested that a passive spectator who is multi-tasking in this way may not be as focused on the information in the tool. This will be taken into consideration when running future studies.

The most active streamer in the study emphasized the need for adding this emotional data to existing Twitch overlays. Most streamers use an overlay that makes use of the available real estate provided on the screen for additional information and metrics related to the game, or that the stream would not normally provide. For example, a time counter for streamers attempting a world record in a game, or advertisements for pro streamers. This participant's observations suggest that the additional elements provided by All the Feels should seamlessly integrate into existing overlay systems while still staying visually appealing to spectators.

Participants had some trouble with the graphical representation of the information on-screen. In particular, some did not understand what the number for GSR meant, and all felt that the data overall could be represented in a clearer way. Participants also remarked that it was hard to see the numbers and the emotional data on-screen. They would have liked to see the data displayed larger and more clearly. For the next phase, we worked with a graphic designer to improve the visualization of the information.

2.2.5 Conclusion and Future Work

We created a prototype of a tool for sharing real-time biometric information in gaming streams. We conducted preliminary tests of the concept using a prototype to engage streamers and spectators. Our results suggest that the tool has promise, in particular for casual streaming audiences, and perhaps also for female audiences. It could be used for player stress regulation, as well as an interesting addition to existing overlays for the spectators. Future work at this stage included representing the emotional data in a more spectator-friendly and intuitive way, and revising the tool to integrate seamlessly with existing streamer workflow, as well as research to better understand how spectators interpret and respond to the data. There was also work to be done in understanding how best to represent these traces of play performances after the fact, building of course upon the experiences of [28].

2.3 Phase 2: Implementation of All the Feels

2.3.1 Design

All the Feels was designed to take advantage of and build upon the customization culture of Twitch. Our previous work with the early prototype developed in Phase 1 [29] provided a good starting point for our choice of sensing technology: something that wasn't disruptive to the player and that provided multiple forms of engaging biometric data.

Next, we considered the choice of sensors and which biometrics to measure. We decided to explore two approaches with different but complementary strengths: 1) making available invisible physiological indicators of emotions, such as electro-dermal activity (EDA) and heart rate, and 2) enhancing already visible emotional cues from the streamer, such as facial expressions of emotions. For the former, we explored different biometric tools that could be used to capture the player's EDA and heart rate. We wanted the system to be unobtrusive and easy to setup and use for the casual gamer, so that the system could be tested in naturalistic settings with and without the researchers physically present. In the second phase of the development of All the Feels, we moved from the Microsoft Band to the Empatica E4 wristband, because it provided us with

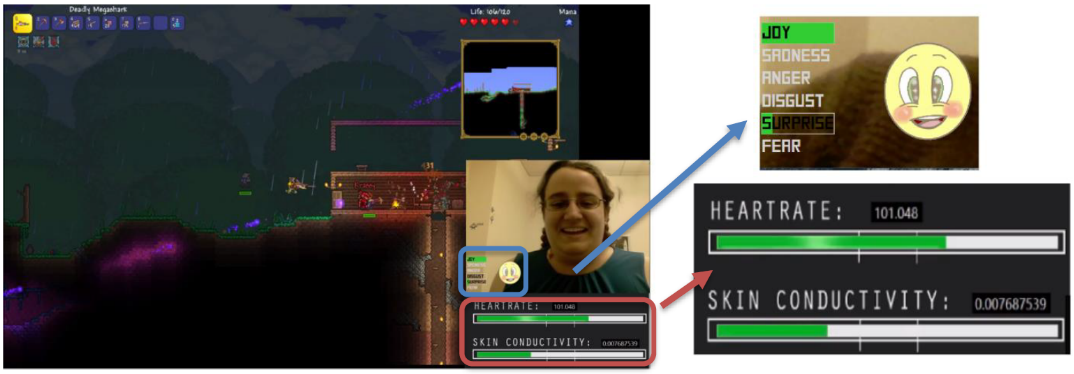

Figure 2.6: Gameplay stream with the addition of All the Feels' and a close-up of each element of the tool. Data from Affdex is marked with the blue arrow and the Empatica data is marked with the red.

more reliable EDA and PPG sensors (PPG stands for Photoplethysmography Sensor and it measures blood volume pulse and heart rate BPM) that could potentially be used to display physiological cues related to the emotional arousal of the player, as well as their engagement, stress, and excitement. In addition, the sensors came integrated in the Empatica, which could communicate wirelessly with a Bluetooth dongle that attached to a computer. This feature addressed the comfort of the player as well as the low requirements and technical skills to set up the ATF tools. Empatica data could be communicated through a mobile application or windows software. We opted for the latter to minimize the number of pieces of hardware necessary for system setup.

Regarding data representation, we opted for displaying continuous data in the form of a bar graph. This was motivated by previous work [29], where potential users expressed their preference for this form of visualization over numerical representations. Hence, we developed software to take EDA and heart rate data from the Empatica and represented them using two continuous bars that were updated real-time with the player's biometrics (Figure 2.6).

We calibrated the bar graphs by evaluating the biometric responses of two graduate students during periods of rest, light physical activity, and heavy physical activity. From these evaluations, we set the middle of the heart rate bar to 80 beats per minute (BPM), and the range from 0-160BPM. The Empatica's EDA measurements decreased during heavier activity, so we represented that data as an inverse relationship. We set the middle of the bar to .005μS and a range of .001-.01μS. As for the recognition of facial expressions of emotions, we used Affdex, which is facial recognition software able to distinguish between six expressions: fear, surprise, disgust, joy, sadness, and anger. We decided to explore two ways of representing this data: a) graphically with one bar graph that continuously showed the real-time perceived level of each particular emotion and b) with emojis that represented the predominant emotion when it reached maximum level in the emotion bar. The choice of emojis is motivated by research that shows that they are not only a widely used and highly enjoyable tool for online communication, but are also a valuable addition to communication methods [10]. While the bars show continuous and subtle changes in facial expressions, the emojis highlight peak facial expressions of emotions.

We used Open Broadcasting Software (OBS) as the live-streaming and video editing software to connect and stream to Twitch. Other software packages (FFSplit, Xsplit) had issues with window transparency for the applications running. Additionally, the other services provided fewer stream customization options than OBS did.

Finally, regarding the overall design for the overlay, we wanted the tool to take up as little screen real estate as possible so as not to interfere with the visualization of the stream, but at the same time enough to allow the spectator to see these measures. Internal in-lab testing showed us that 25% of screen real estate was a good compromise and that the lower-right portion of the screen was the best position at which to display the tool, because that space provided the least useful information in the games played.

2.3.2 Methodology

We ran a preliminary study to explore the potential of ATF to enrich the spectators' experience, looking in particular at their engagement, enjoyment, and emotional connection with the player-streamer. The streamer we recruited for this study is a PhD student at the university where the study was conducted. She mostly streams to Twitch creative and has about ten or so viewers that regularly watch her streams. For this study, we recruited spectators through advertisement: posting on Facebook and Reddit as well as sending out emails requesting the link to the stream be shared. The study was conducted in the streamer's lab on a Friday at 7:00PM PST. The time was chosen since it's considered peak internet use in several contiguous time zones to ours. We were physically present in the same room as the streamer to setup the tool and assist in the event of technical difficulties.

The stream lasted for two hours. In the first hour, she streamed normally while playing Terraria, an action/adventure sandbox video game. After that, we set up the ATF tool and she continued her streaming of Terraria. After 15 minutes, she switched over to play Five Nights at Freddy's: Sister Location, a point-and-click survival horror game. Both games were chosen by the streamer, although for the second game, we instructed her to choose a game that she thought would generate a high physiological response. The entire live-stream was video recorded and saved for later analysis. The streamer played at a laptop, and alongside her was a second monitor displaying the stream, ATF, and Twitch chat comments. Due to internet speeds, there was about a 40 second delay in the live-stream. The way we set up the monitors, the streamer would have access to her emotions, as well as the comments by the spectators. We expected that exposing her biometrics would stimulate the chat conversations, and trigger comments about the spectators' perception of her emotional state. The streamer could in turn respond to this, verbally and non-verbally, and it could also influence her emotional state and gameplay.

At the end of each hour, a survey was sent over the chat to the spectators to understand more about their experience while watching with and without ATF, and their opinion on the tool. Survey measures were drawn from related works [21, 25] and adapted to fit this study. In particular, we asked spectators about their enjoyment of the tool, and about their engagement as reflected by how frequently they checked the ATF and how long they watched the stream. We also asked them to rate their feeling of connection to the streamer. In addition to the survey, we analyzed interactions in the chat to know more about spectator enjoyment and engagement. In particular, we noted the amount and frequency of participation of spectators, and comments and discussions about the biometrics of the player as well as events in the gameplay.

Finally, we conducted an interview with the streamer in which we replayed the stream back to her, using a retrospective think-aloud approach to gather her opinions about her stream experience and ATF. We asked her about her enjoyment of ATF as well as her perception of connection and social engagement with spectators.

2.3.3 Results

Spectators were asked to fill out a different survey at the end of each hour to compare their experience of the stream with and without the addition of ATF. Due to limitations of the Twitch platform at the time, we were not able to determine spectator count at each moment of play. During the first hour of streaming we had nine spectators fill out the survey. We experienced an increase in survey participation to 17 individuals during stream hour two, when biometric data (ATF) was added. Six of the 17 spectators that filled out the second survey during hour two also watched the first hour and filled out the related survey.

In the following, we present results from our analysis of the video recording of the stream, chat log data, and survey responses. Results are presented first as regards spectator enjoyment of ATF and next as regards spectators' perceived connection to the streamer and overall engagement. We used a mixed-methods approach to analyze the data collected. In the future, we plan to conduct a study with a larger number of participants, allowing for a more in-depth quantitative analysis.

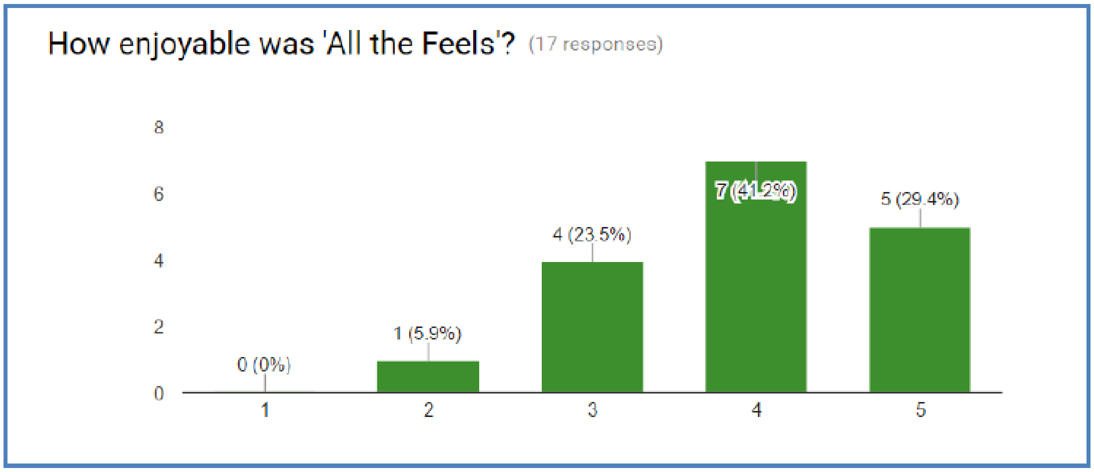

Figure 2.7: Spectator enjoyment of ATF on a scale 1=Not enjoyable; 5=Very enjoyable.

2.3.4 Enjoyment Ratings of ATF

At the end of the streaming session, we asked spectators to rate their enjoyment level on a scale from 1 (Not enjoyable) to 5 (Very enjoyable). Figure 2.7 shows the spectator enjoyment of ATF as reported by the spectator during the post-stream survey. The mean enjoyment was 3.94 ($\sigma = 0.9$), indicating that the majority of users found the tool as enjoyable.

Spectators stated that they looked at ATF frequently and reported that they viewed it quite often ($Mean = 4.29$; $\sigma = 0.59$) (Figure 2.8). Spectators were also asked to rate the degree to which the tool was distracting using a scale of 1 (Not distracting) to 5 (Very distracting). The percentage of spectators that found ATF distracting to some degree was 82.3%, though degree of distraction was variable (Mean=2.65 $\sigma = 1.17$). Four spectators suggested that reducing the real estate taken up by the overlay would reduce the distraction factor the most. Participants also indicated some interest in

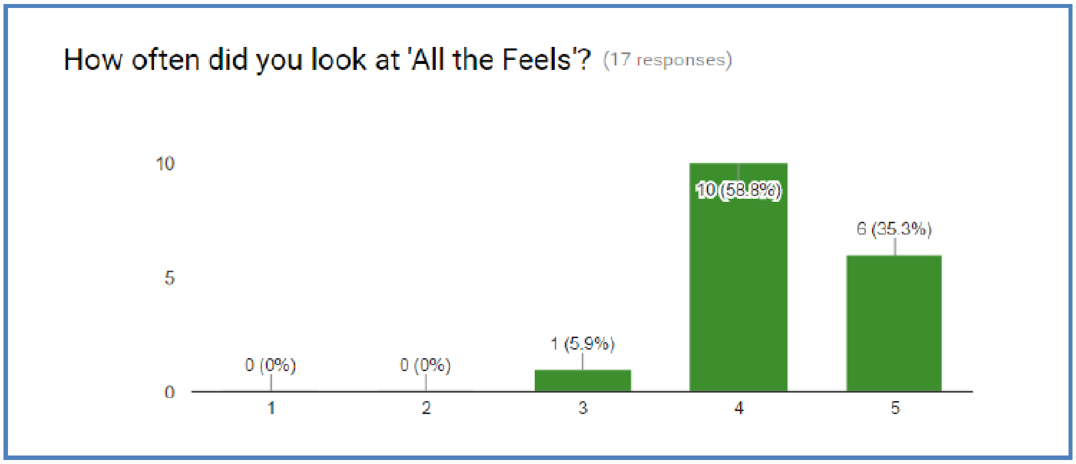

Figure 2.8: Spectator viewership of ATF on a scale 1=Did not view at all; 5=Viewed very often.

seeing ATF in other Twitch streams ($Mean = 3.77$ $\sigma = 1.03$). Spectators were asked which metric provided by ATF they looked at the most. Eight participants looked at more than one metric with five preferring HR and emoji, and three preferring HR and EDA (Figure 2.9). No participants looked at EDA by itself. The question regarding opinion about the representation of facial expressions of emotions did not distinguish between the emoji and the six emotion bars. It is unclear whether spectators enjoyed one or both aspects, however two participants further clarified this in a final open question, indicating that they looked at the emotion bars the most. Five females and four males tuned in to the stream without the overlay. Six spectators watched greater than ten minutes and three watched the entire stream (33.3%). For the ATF stream, 13 males and four females watched. Eight watched greater than ten minutes, six watched the entire stream (35.3%), and three watched 5-10 minutes. ATF produced higher

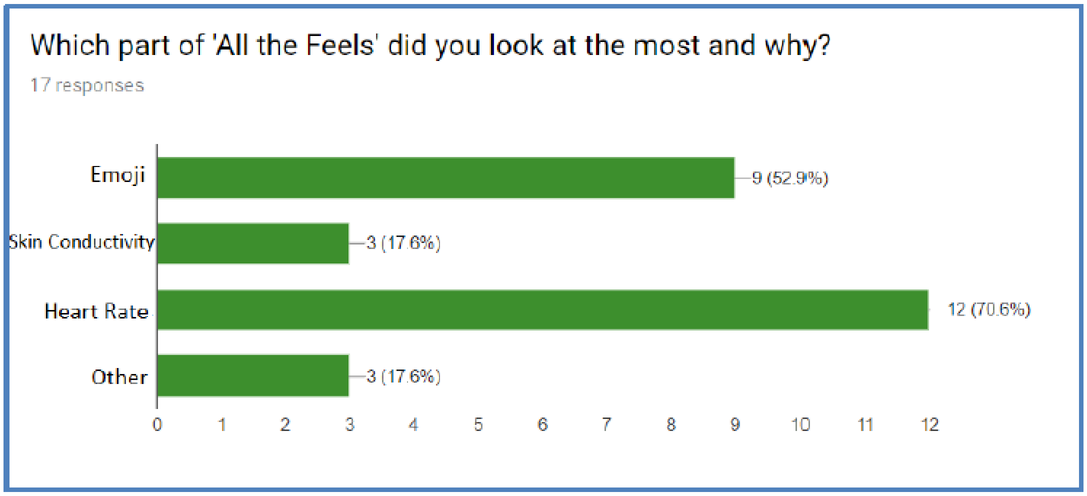

Figure 2.9: Spectator viewership of ATF categorized into different parts: emoji, skin conductivity, heart rate, and other.

consistent viewership time among a higher number of viewers.

In the post-stream interview with the streamer, she indicated low distraction and no discomfort while using ATF. She was undecided about whether she would have liked to see ATF readouts more clearly or not, since she was worried that a clearer view would increase distraction levels for her: "I would think, can I make facial expressions that generate those emotions?" Accuracy of the data was important to her, and she would be willing to sacrifice comfort for performance. While the streamer could look at her emotions on the second monitor, she also experienced no discomfort in exposing her emotions to spectators or being able to see them herself. She felt that spectator comments helped her remain calm through the horror game and made her feel less alone. She noted this as a sympathetic response from the spectators. Spectators were overall very encouraging and helpful toward her through the length of the stream, providing words of comfort and encouragement during peak moments of play.

We aimed to investigate spectator engagement through the frequency of conversation in Twitch chat logs and perceived connection to the streamer. Not only did spectator count double in the ATF stream, chat participation doubled as well with an increase of 1.08 comments per minute to 2.08 comments per minute. Total comments increased from 56 comments for the non-biometric stream to 133 comments for the ATF stream. About 5% of total comments during the ATF were about the data provided by the overlay. In the non-biometric stream, there were 5 spectators commenting on the stream and 4 silent. Of the users interacting with this stream, the average number of comments per user was 9.2. In the ATF stream, there were 9 spectators commenting on the stream with 8 silent. Average number of comments among active users was 13.45. The average comments among active users increased from the non-biometric stream to the ATF stream, which indicates an increased engagement among the spectators. Spectators commented on the behavior of the emotion bars and the heart rate meter during high stress periods of the games. Multiple quotes from participants including "the fear and sadness is capping a lot HAHA" and "heartrate shot up so much oh my god" contain remarks about the streamer's biometrics, suggesting a close monitoring of the streamer's reactions and emotional state and ultimately an interesting emotional connection to the streamer.

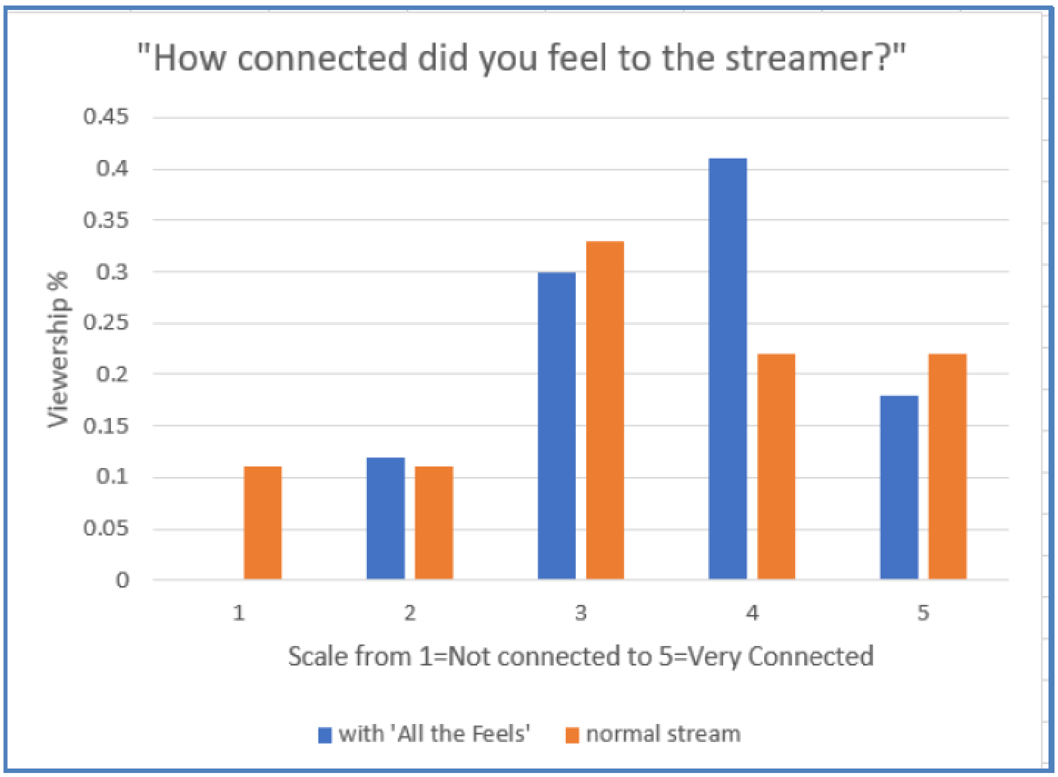

Figure 2.10: Perceived connection to streamer by the spectators.

2.3.5 Spectator Engagement

Figure 2.10 shows the percentage of viewers that felt connected to the streamer on a scale from 1 (Not Connected) to 5 (Very connected). Percentages were used due to the difference in number of spectators between the normal stream and ATF. Average connection for the normal stream was 3.33 ($\sigma = 1.32$). Average connection for the ATF stream increased to 3.64 ($\sigma = 0.93$), suggesting that spectators may have experienced higher and more consistent degrees of connectedness. (It is important to note that with the small number of participants we were not able to run significance analysis on the results, so more research is certainly needed.)

2.3.6 Discussion

Displaying performers' physiological data has increased engagement and empathy of spectators in other contexts discussed earlier [34, 37, 38, 39]. Preliminary results of the study we conducted align with this prior work done, showing increased enjoyment, engagement, and viewership among live-stream spectators.

We found that ATF impacted spectator experience the most when the game produced large changes in the physiological responses of the player. Spectators liked either the heart rate or Affdex emotion labeling data the most. Spectators also indicated higher enjoyment levels of the ATF stream, and there seemed to be a trend toward increased connection to the streamer. The number of spectators and the number of comments by the spectators increased from the non-biometric stream to the ATF stream, showing a heightened engagement with the stream. Additionally, the tool was not disruptive to the player in any way.

We received valuable feedback in survey responses from spectators and we plan to continue to investigate optimal strategies for improvements in design. We would also like to note that this stream was a particularly friendly and supportive community, which may not always be the case in future studies and streams. In this case, we accounted for this issue with moderators tasked to watch the stream and make sure not to allow harassment or toxicity.

2.3.7 Limitations

The test was conducted in a laboratory environment. The streamer reported that she streams the most in her bedroom and is most comfortable there. This study would benefit from a more natural testing environment, something we can explore in future iterations. Additionally, we were physically present in the room to moderate the stream and help with any technical issues that might arise. More than once the streamer interacted directly with us outside of Twitch's social options, unintentionally interacting in ways outside the scope of the study. And of course, any given single streamer cannot represent the potential range of response of all streamers.

While all software for ATF worked flawlessly, the hardware did not. The Empatica had some accuracy and latency issues. Heart rate information was found to be accurate, however initial HR acquisition took 1 minute and 30 seconds. Additionally, the heart rate was slow to update. While it worked for the three-main jump-scares that occurred in Five Nights at Freddy's, it was variable during less active moments of play. However, even with the accuracy issues spectators still showed interest in the tool and enjoyed the experience.

2.3.8 Conclusion

We have presented a tool that integrates biometric streams and emotion recognition data into Twitch video game streams, toward improving the spectator experience, and feelings of connection with the streamer. We conducted a live test with the tool involving a female streamer and 25 spectators. We found that All the Feels seemed to increase perceived connection between the spectator and the streamer, while offering an engaging and enjoyable experience for spectators. The player did not find the software obtrusive, and appreciated the supportive dialog that ensued from spectators who saw her emotion data. We believe this research shows that All the Feels is a promising addition to gameplay streams, to be proven out in more in-depth, extended field testing. The results of this research support our concept that biometric data could be used as an additional layer of information to spark conversation and enhance streamer and spectator experience. We plan to continue to evolve the design of All the Feels based on the feedback from this study, and to test it with a broader range of streamers in future.

Chapter 3

Player Experience

3.1 Background

In 2008, a CHI workshop on "Measuring Affect In HCI: going beyond the individual" recognized that while techniques for measuring affect in HCI have been developed, a systematic comparison of effectiveness and applicability in different contexts was lacking [34]. Since then, various comparisons of user experience methods have been conducted [38]. However, no comparison of methods measuring players' emotions in games has been conducted to date.

Research in these other domains of HCI supports using mixed modalities to capture a range of users' emotions [11, 38]. We would argue that this is especially important in games, due to the more nuanced range of emotions that games give us [11]. Games are functionally different from other interactive systems such as bank teller machines or web browsers because they have different operative goals. Their focus is on entertainment, whereas the primary goal of the latter is function. Games give players the chance to influence outcomes through their own efforts [32]. They can elicit pride, guilt, and complicity – emotions that are uniquely experienced in games as compared to other media [11]. To evaluate these varied emotional experiences, a variety of techniques have been developed and analyzed over the years.

Questionnaires are the most commonly used of all the techniques, and have been developed to measure the player's overall experience, immersion, satisfaction, attribution, and so on [8, 7]. The Player Experience of Need Satisfaction questionnaire (PENS) and the Game Experience Questionnaire (GEQ) are considered to be the most widely used for evaluating how players are feeling about a game [10]. These questionnaires use Likert scale structured questions, varying positive and negative construction. For example: "I felt that I really empathized with the game." followed by "I did not feel any emotional attachment to the game." In 2015, Bruhlman conducted an in-depth analysis of the PENS and GEQ [5] and in 2016, Denisova presented a comparison and analysis of the numerous existing player experience questionnaires [7], and suggestions for converging and improving them.

Another method is post-interviews, in which players are asked after the experience is completed to answer a variety of questions about gameplay and describe how they felt at certain moments of the game. In these cases, the video footage of their playthrough is shown to the player and they are asked to recall and describe how they were feeling at specific moments of play. However, this technique can be limiting because of the difficulty of recalling emotions felt during the game after the fact. Therefore, this technique is usually combined with in-game think-aloud. Think-aloud is a method to determine how the player feels about moment-to-moment play. With this technique, the player talks through decisions and choices they make during the game itself. A limitation of this method is that gameplay can be quite cognitively demanding, and it may not be possible to comment upon the experience thoroughly while staying immersed in play.

Most of the techniques described above, in particular the questionnaires, have been designed to enable gathering emotional response from a large number of players. They are not tuned to capture all the nuances and complexities of emotions in each individual player's experience. What if we want to understand the player's rich responses to the twists and turns of an interactive story, for example? Dewey states that emotion is an important aspect of all storytelling experiences, one that binds an experience into a whole [9]. He notes that emotions are not static but change over time, together with the experience. There are additional methods have been developed with the goal of surfacing the kinds of complex, intermingled, less visible emotions that one might expect to arise from an interactive story. One such method is the Sensual Evaluation Instrument, or SEI as we will refer to it throughout the paper [12, 18]. This is a set of eight objects (see Figure 3.1) that range from spiky to smooth which are placed in front of the players for them to use during the game.

Rather than have a 1-to-1 mapping with any particular emotion, the SEI are left open ended, for the player to decide which emotion maps to which object(s). When SEI objects are used during game play they function as rich conversation markers that

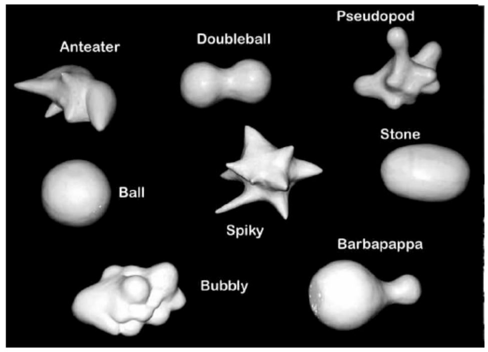

Figure 3.1: Sensual Evaluation Instruments.

open a discussion about players' emotional experiences, between the designer and the player. Before usage, the SEI must be calibrated, where 15 (positive/negative) images are shown to the player and they are instructed to pick up one of the SEI objects. This grounds each object to a potentially related experience. Other instruments like the Subtle Stone [3] were developed; similar to the SEI, acting as a tangible tool for communicating emotion in a classroom environment. The SEI was created in 2006, and in 2009 a paper was written discussing the applications and usefulness of the tool in various narrative games [18].

Physiological data gives us what some might characterize as objective insight into the player experience [21, 22], because it is grounded in signals measured from the body of the person having the experience. Capturing players' physiological data is a valuable method in the games domain, because it can give designers access to signals of player emotions that may not rise to the level of active communication, and because these signals can be gathered over time readily, and linked to designed game events.

Figure 3.2: Empatica E4 wristband. Captures a variety of biometrics, including heart rate and electrodermal activity.

For example, storyboards have been created which visualize meaningful relationships between a player's physiological changes and game events [23]. Electrodermal activity (EDA), which is skin perspiration or arousal, has also been used to evaluate player experience [14]. However, the physiological sensors used in prior related work have been fairly bulky and intrusive, which could potentially affect the player's comfort, and therefore immersion levels. More recent tools, which strike a balance between accuracy and intrusiveness, have been developed. Our tool, All the Feels, [31] is a less-intrusive way to capture players' physiological data in games. This version of the tool utilizes an off-the shelf wearable device (see Figure 3.2) and facial recognition software (see Figure 3.3). The wearable is called the Empatica E4 wristband and captures heart rate and Electrodermal activity (EDA) or sweat as it is labeled with the tool, from the player.

These metrics are displayed as a continuously updated side bar graph (Figure 3.3). The facial recognition software is called Affdex, which captures and displays players' micro-expressions as individual bar graphs: joy, sadness, anger, surprise, disgust,

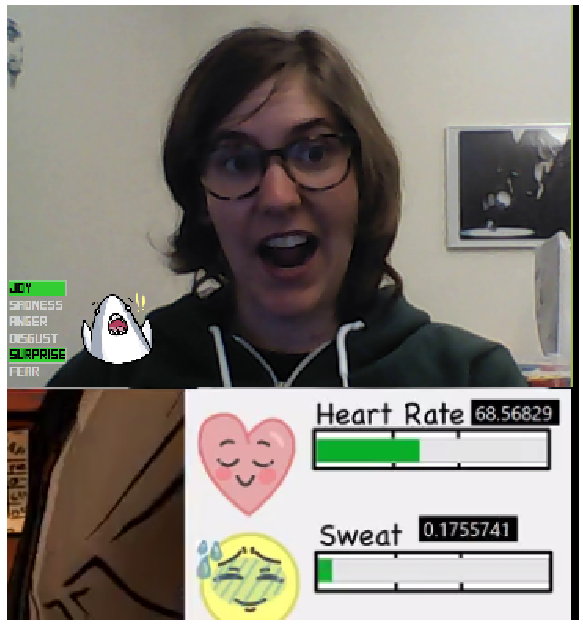

Figure 3.3: All the Feels. Used in the study to capture facial expressions, heart rate, and sweat.

and fear. When one of the 6 emotions reaches 100%, an emoji is displayed corresponding to that particular emotion (Figure 3.3). Facial expression analysis has also been used to infer player experience in numerous other contexts [26, 36].

In designing our study, we chose a set of emotion evaluation methods drawn from those discussed above. We planned to use both verbal and non-verbal techniques, as emotional expression takes place heavily through nonverbal cues as well as verbal ones [16]. While verbal techniques can give a lot of insight into what a player is feeling and their thoughts about a game, they leave out non-verbal expression, which also seems important for attaining insight into how players are feeling.

We chose think-aloud and post-interviews and questionnaires as the verbal methods. This combines thoughts on moment-to-moment play with the reflective perspective that post-questionnaires and post-interviews provide. We chose the SEI and All the Feels to capture nonverbal emotional response. The project team was interested in the interactive storytelling genre because of ongoing research, and because players can influence the progression and outcome of a story within this genre, providing for a more dynamic experience than traditional stories [17, 19, 24]. This dynamic experience, we felt, would allow us to see a range of emotional responses throughout gameplay. Within the genre, we chose The Wolf Among Us, a narrative-focused game that requires players to make a series of choices. The plot of the game focuses on an investigation by the player-character protagonist, Bigby Wolf, of a murder in the community of Fabletown, where creatures and characters from bedtime stories have been forced to flee their homelands and take up residence in New York City.

3.2 Methodology

The overall study design was to have several people play through the selected game, using all of the emotional evaluation methodologies in parallel. Because of the in-depth comparative coding and analysis we planned to do, we limited this study to 7 participants, 3 female and 4 male, with an age range of 20-35. We chose participants who had never played the game used in the study before. The study was held in a user testing lab at the researchers' University. The modes used in the study by the players were think-aloud, retrospective interviews, a tailored version of the IEQ, the Sensual Evaluation Instrument, and All the Feels, following recommended procedure for each

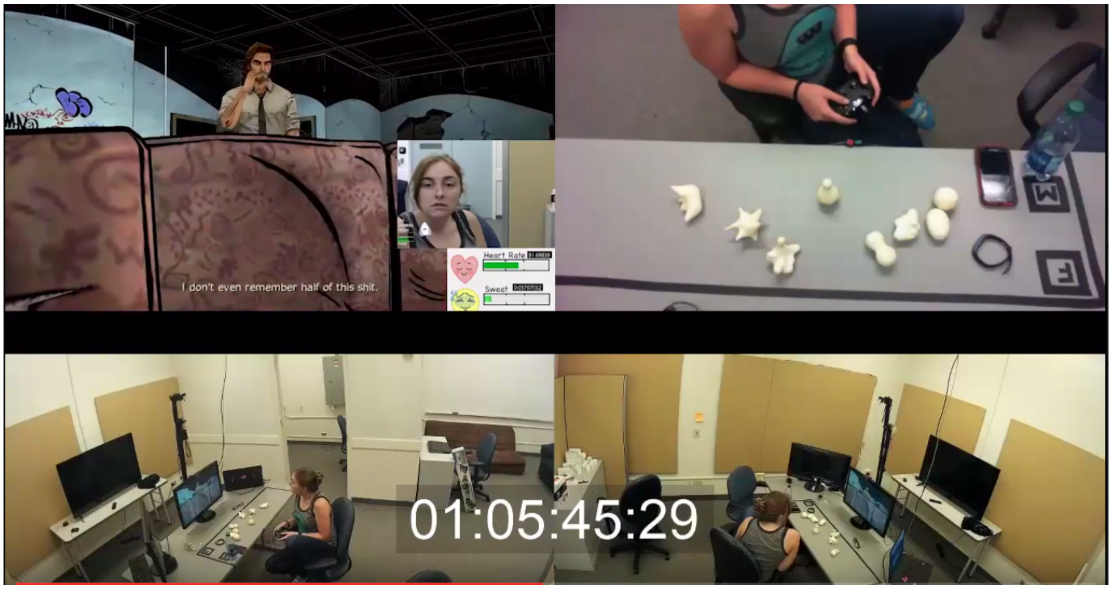

Figure 3.4: Lab setup including biometrics, SEI, and 4 cameras. Upper left is front-facing webcam, upper right is top-down webcam focused on the SEI, and lower left and right parts are the left and right room webcams respectively.

instrument [11, 31].

The lab was set up with two cameras on each side of the room to view the left and right angles of the player, a top-down camera focused on the SEI, and a front-facing webcam capture of the player's face to use for the facial recognition software (see Figure 3.4). The SEI objects were laid out in front of the users, and they were instructed to use or configure the objects throughout the course of the game. A console controller was used in place of a mouse and keyboard to make it easier to use the SEI during the session. We set an alarm to go off every 15 minutes to remind participants to use the SEI, because we noticed in a pilot session that the player did not remember the SEI were there when she became more immersed in the plot. For All the Feels, the players were equipped with the Empatica E4 wristband to capture the biometric data: heart rate and sweat. However, this data was not shown to the players; all they could see was the game and the SEI laid out in front of them.

We began each session by having players take a pre-game questionnaire, asking them if they were familiar with the Fables comics the game is based on, or any other games by the same company that might impact their feelings toward the game. We calibrated the SEI, which involves showing the users a set of 10 images drawn from the International Affective Picture Set and having them indicate which SEI object they associate with which image.

We then instructed participants to play the game, think aloud, as well as use the SEI whenever they felt like it. All the Feels was running in the background, capturing players' physiological data throughout the course of the game. At the end of the play session, we conducted a retrospective think-aloud with the players and asked them to talk about how they were feeling at specific moments of the game. We also asked them general questions as to how they felt about certain characters, and overall gameplay. We had them take a post-questionnaire, which was a tailored version of the IEQ discussed earlier. This questionnaire had both positive and negative general questions about the game, along with more specific questions about immersion, affect, and players felt about specific characters and events in the game.

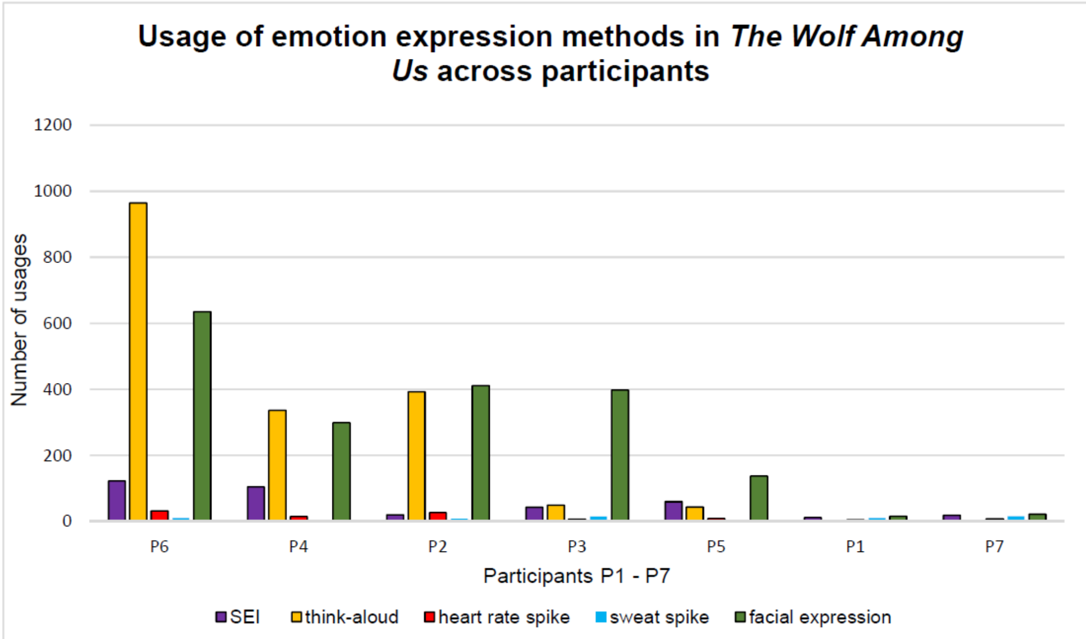

Figure 3.5: Graph showing the usage of each method across participants: SEI, think-aloud, heart rate spike, sweat spike, and facial expression.

3.3 Results

We used Adobe Premier Pro to code the video files of each participant's session. We set a marker on the video each time a participant used a SEI object, said something as part of the think-aloud protocol, made a facial expression, and experienced a biometric spike (see Figure 3.5). We classified think-aloud as each independent expressed thought – a sentence, word or phrase. Heart rate spikes were marked as anything above 10BPM from what the player's average heart rate was throughout the game. Similarly, with sweat, a spike was categorized as anything over 1-2μS from their average. Facial expressions were recorded whenever the facial recognition software displayed an emoji (i.e. whenever one of the 6 emotions hit the threshold within the Affdex software for strong emotion signaling). We have labeled the participants P1-P7. To assess the methods, we used both qualitative and quantitative strategies, but due to time constraints, we are only able to report certain elements of our analysis for this late-breaking work. In the future we plan to produce a more in-depth look at this large data set, and report on our findings.

One striking finding of the study was how much individual variance there was in how players used the methods. The players had a wide range of expressivity, shown in Figure 3.5, with the left-hand side of the graph showing the most expressive of the players and the right-hand side the least. We categorize expressivity in terms of high usage, moderate usage, and low usage of the modalities. P4, P2, and P6 had similarly high usage of both facial gestures and verbal think-aloud. P3 and P5 had similar levels of expressivity: moderate usage of the SEI and think-aloud, with facial expressions as the most used method. Both P1 and P7 had low levels of expressivity across all methods. To illustrate this range even further, we will provide an example, which we will label Example 1, from a scene in the game. In this scene, the player-character Bigby arrives home and finds a pig named Colin asleep in his chair. Upon meeting Colin, the player engages in a conversation with him, and is presented with two decisions to make – a) whether to kick him out of the chair and b) whether to give him a drink. Figure 3.6 illustrates the usage timeline for each player. The left column indicates the player number, and the right column includes the timeline with the markers depicted by color. The yellow marker indicates that the player thought aloud, red a heart rate spike, green a facial expression, blue an EDA spike, and white is use of the SEI. The bottom of the

Figure 3.6: Left column indicates player number and right column shows usage timeline. Key describes which color markers correspond with which methods. The entire scene lasts about 5 minutes and 30 seconds, which can be seen above the key.

image shows that the scene lasted about 5-and-a-half minutes. This figure aligns with the data presented in Figure 3.5, showing further evidence of this variance in player expressivity at a more fine-grained level. However, despite these personal differences, we still observed similar patterns of usage among players.

Each player experienced a similar number of biometric spikes — about 10-20 moments during the game. At peak moments, biometrics were useful for observing the player's heart rate jump by 10 beats per minute. Moments that caused these spikes could be clustered into two categories: 1. QuickTime events – i.e. moments of the game that force the player to click certain buttons in a series to advance, and 2. Tense or suspenseful moments of the game. The bottom of Figure 3.6 shows the timestamps of the scene from Example 1. As you can see at 4:30, four out of seven players experienced a heart rate spike at the same moment in the scene. SEI usage varied amongst players, depending on their levels of expressivity. Usage sometimes referred to what was happening on screen, a reflection on the story state, while other times it referred to the player's own experience of the story. Players also used the SEI if there was not an option for action to take that they liked on screen, or if they wanted to express how they felt about a decision they had just made. The ways in which players used the SEI varied as well. Both P4 and P6 grouped the SEI objects into a circle and put the one they most wanted to express in the middle of the circle. P3 would often put one of the objects in their lap while playing, so they could keep their hands on the game controller while simultaneously expressing their emotions with the SEI. P2 would touch the SEI and accompany the touch with a corresponding think-aloud of the specific emotion that went along with it.

To provide another example demonstrating alignment among SEI usage, we will focus Example 2 on a scene where the characters are investigating Prince Lawrence's house because he was suspected to be dead. The scene is tense and the music suspenseful. Bigby walks up to the closet door preparing to open it. The music intensifies; tension building. Just as he opens it, an intruder barges out and runs away, transitioning the game into a chase scene. This is a point in which a tense scene has ended — and now a QuickTime event has begun. During this transition, 5 out of 7 participants reached for the spiky SEI, the object in the set typically associated with anxiety (see Figure 3.7). One participant used the SEI earlier in the scene. In each case, the SEI was used when the intruder character named Dum appeared on screen. Each participant, when interviewed about how they felt during this scene, indicated some amount of anxiety, tension, or buildup in the game that caused a negative feeling.

To provide an example that illustrates the value of think-aloud, we will focus

Figure 3.7: P2-P7 using the Spiky SEI during the same scene in the game. The red circle shows the spiky object.

Example 3, on an alignment among thoughts expressed by P4 and P6. Both participants had the same comment about two characters in the game. When Snow was first introduced, P4 said "Why does everyone in this game look like Elizabeth from BioShock?" and when P6 saw Faith for the first time, they stated "Ohhhh she is Elizabeth from BioShock Infinite." When each of them was asked about this connection, they both stated that both Faith and Snow looked like Elizabeth. This response to the game would not have arisen using any of the other methods for gleaning emotional response to the game — it was a fleeting moment of recognition that led to contextualizing the game among other related games, which shaped players' emotional responses to the game. From a design point of view, this response is interesting and useful in understanding players' overall emotional reactions to the game.

Example 4 focuses on facial expressions. This is a scene toward the beginning of the game, in which Bigby gets into a fight with the Woodsman, who at the time seems like a villain that is beating up a woman. In this scene, there is a moment where the player's character is being choked by the Woodsman. During this scene, P6, P2, and P4 (top and middle rows of Figure 3.8) made similarly concentrated facial expressions, which seem to combine pleasure with effort. These all registered in the facial expression analysis software as joy. The last row of the figure shows P1 and P7 at the least expressive, with no discernible facial expression. (Note: The physiological data is calibrated and based on the players' own baseline data. Therefore, while two players' BPM may appear to be the same, they are different for each player depending on the base level. The middle of each heart rate and sweat bar represents their "baseline

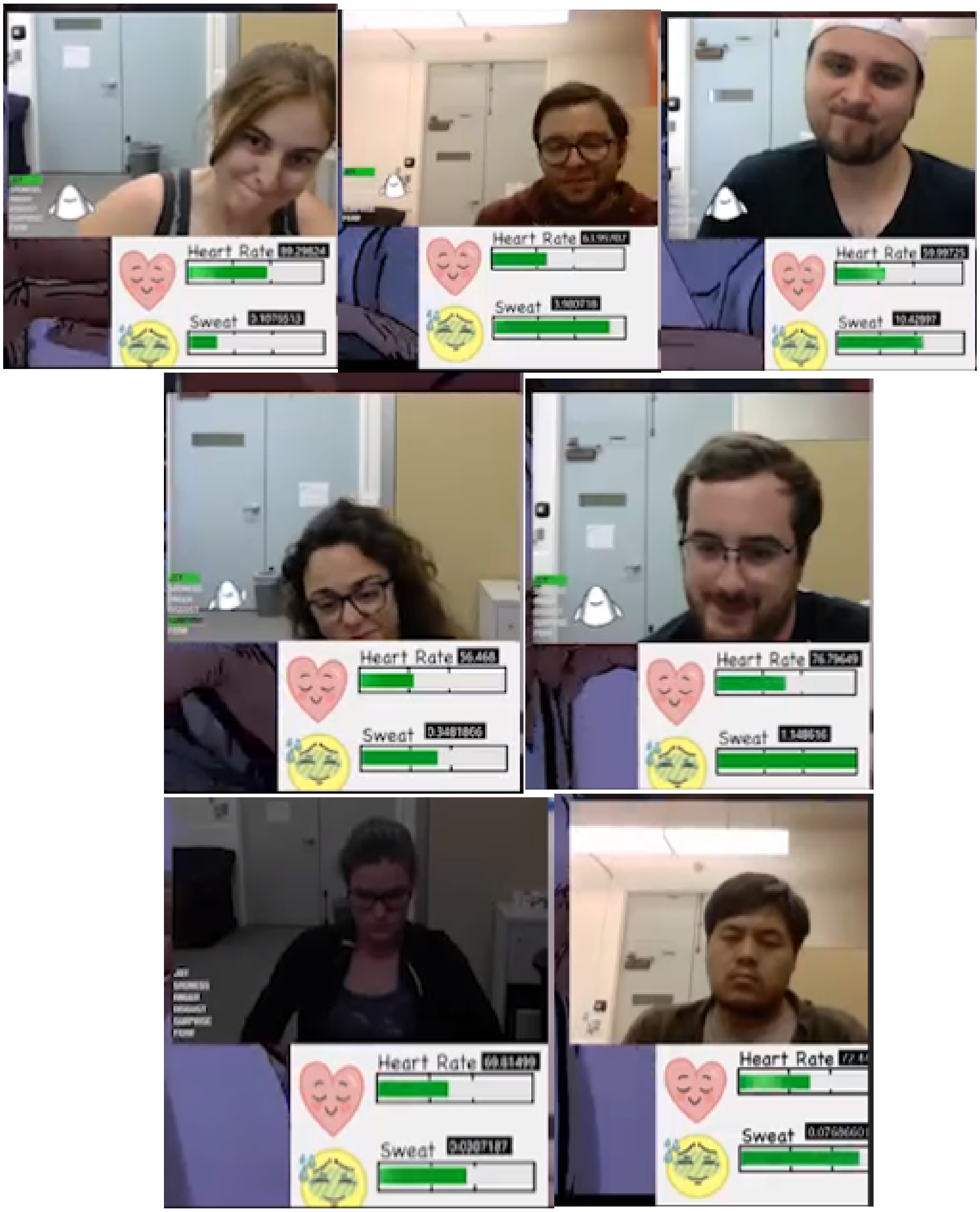

Figure 3.8: Facial expressions and physiological data of all participants during the first fight scene.

BPM".)

The post-questionnaire consisted of 39 items asked players both negative and positive questions about immersion, enjoyment, how the controls felt, and agency. We used 25 items from the IEQ [13] and the rest asking game-specific questions. Questions were on a Likert scale from strongly disagree to strongly agree with the statement. We omitted IEQ questions about winning/losing the game as well as ones regarding how much players enjoyed the aesthetics. Due to the nature of the study and the narrative genre, there was no competitive element, and thus we did not feel this information would provide us any additional insights. Also, we were not concerned with how players generally liked the aesthetics. We replaced these questions with ones about how players felt regarding specific in-game characters. We wanted to know if there was any relationship between participants' usage group and their indicated level of immersion. P1 (one of the least expressive participants) rated an 8 out of 10 in immersion but was in the low-usage group of the methods. Similarly, P3 rated an 8 in immersion, but was in the moderate-usage group. Additionally, P2 rated a 6 on the same question, but was in the high-usage group.

The retrospective post- interviews focused on having the players discuss game and method specifics. We conducted these post-interviews by having players look at video footage of their playthroughs, while being asked questions about their emotions during the corresponding scene. We first asked them to recall any key moments where they used a SEI instrument and what was the scene in which they used it. We asked them to describe the scene, why they chose to use that object, and what emotion they were feeling at the time of usage. We also asked if there were moments in which they felt like they should have picked up a SEI but were too distracted, and most players indicated there were not any scenes in which they felt that way. We also asked them how they felt about game-specific scenes such as Beauty hiding her presence from Beast, and how they felt about Crane and Grendel as characters. During the interviews, they tended to talk through their emotions about these scenes and indicate what they were feeling. And, if they used a SEI during that scene while playing the game, they would reach for the SEI during these interviews as well. Overall, players felt similarly about key moments of the game, and indicated similar emotions toward the characters and events we asked them about.

3.4 Discussion

In this section, we discuss the results in terms of the examples provided in the previous section: Examples 1-4. We hope to impart a more nuanced understanding of this suite of methods – and give some insights as to the variations in the way players' emotions manifest using this suite of methods and which methods are useful at which moments of play. In Example 1, we see a high variance in the usage of the methods among players, which speaks to the expressivity of the individual. This could be because of personality differences, or simply differences in the way people react to media. In the future, we may deploy a personality test to help better understand these differences, and to see if there is any relationship between expressiveness in responding to a game and expressiveness in real life. We would also like to note that players' performance abilities may influence their expressivity. For example, how used to streaming they are, or how familiar they are with gaming. Examples 2 and 4 show the alignment of the methods across players. The high usage group favored think-aloud and facial expressions. The moderate usage group favored facial expressions. This suggests that each group similarly favors expressing themselves in particular ways. The high usage group, in a verbal way, and the moderate usage group, in a more bodily way. One thought we had is that these differences could be related to findings in the literature on learning styles about peoples' preferences for taking in information — perhaps people also have preferred methods of sharing emotional information [6].

Another interesting point of note in Figure 3.5 is that while facial expressions are captured within the physiological data tool we used – and are generally considered to be physiological, they should not be grouped with biometrics in this context. The facial expressions get used analogously to think-aloud, rather than biometrics. Players use facial expressions as a supplement (for high usage group) or substitute (for moderate usage group) to think-aloud sharing of emotion. We can see these facial expressions in this context, then, primarily as active social communication, rather than as involuntary signaling of the body's emotional response. The range of methods used help accommodate differences in the way players' emotions manifest. For example, the think-aloud that was illustrated in Example 3. The observation of P4 and P6 making similar Bioshock references demonstrate something that could have only been expressed using think-aloud, giving a glimpse into the players' ongoing internal monologue to reveal information that inflects emotional response.

Biometrics helped to track strong emotions at specific moments of play. Throughout the play sessions, at tense or peak moments, players experienced a spike in their heart rate and EDA. This suggests that a jump-scare moment in a horror game could be a good place to rely on biometrics to track effectiveness of design. In contrast, the SEI objects were used more reflectively, expressing a span of time or a scene rather than a moment. The results section discusses some of the various ways in which players use the SEI. For example, in Example 2, we saw a convergence in player usage of one particular SEI object during the same scene of the game. Players used it at different moments in the scene, but they all converged on the spiky object, known to indicate anxiety and tension from prior studies and through the calibration part of our study as well. The post-game questionnaires helped us to understand players' overall feelings toward the game. They gave players a chance to reflect on the overall experience. In our analysis, we found that these ratings reflected the general experience of play, though of course not capturing the ebb and flow of fine-grained emotions.

Retrospective post-interviews were useful for tying in-game events back to how players felt during moment-to-moment play. Generally, looking at footage from the play through gave the players a chance to reflect on what occurred in the game — which we found useful for supplementing as well as validating the other methods.

The function of the post-game questionnaire and retrospective interviews was similar — they allowed players to reflect on how they were feeling during the game. As was covered at the end of the Results section, we might have expected to see more emotional feedback from players who later reported that they were more immersed in the overall experience, but we did not. This suggests that the variance in players' use of feedback tools like think-aloud and the SEI may have had more to do with individual differences in personality than in their interest in the experience itself.

Overall, in this study, using an array of both non-verbal and verbal methods allowed us to construct a more general picture of the emotional landscape in the less expressive players, and at the same time have a more nuanced and clear picture of the more expressive players. A diverse set of self-reporting tools gave players a more nuanced vocabulary to express their feelings and emotions about the game.

3.4.1 Limitations

During the two-hour play session, there were a few technical difficulties and charging issues with the Empatica E4 wristband. Therefore, we were not able to capture physiological data from certain parts of a few players' sessions. Overall, since biometrics were mainly useful at key moments of play, there was little impact on the data. There were also a few interruptions during two sessions which were not coded in the data.

3.5 Conclusion and Future Work

We conducted a study that included multiple methods for evaluating players' emotional responses to a game. We found that these methods captured different aspects of players' emotional experience. When used together, they capture a larger picture of emotional variance among players. Our results showed differences in individual expressivity, for example some people who were not very expressive verbally and nonverbally, for whom biometrics were nonetheless able to pick up peak emotional responses. Think aloud gave us valuable contextual insights, while retrospective interviews and questionnaires help scaffold the other methods. The SEI was good at picking up emotional responses to experiences that spanned many moments, helping us understand the ebb and flow of emotion, while biometrics were good for signaling key high-emotion moments of play. We hope to impart upon game researchers the idea that when evaluating player experience, a mix of methods in any given study should be considered. In the future, we plan to run a larger study with more participants and to include some measure of personality or expressivity to help us understand the individual differences we saw in the use of the various self-report methods. If we can successfully categorize a few stable player types, it may be possible to determine ahead of time which methods will be most useful for differently expressive players, which could better inform decisions about which methods to use when assessing player experience.

Chapter 4

Conclusion and Future Work

I created a tool for sharing real-time biometric information, called All the Feels. I conducted preliminary tests of the concept, using a prototype to engage streamers and spectators, showing the efficacy and applicability of All the Feels in multiple contexts.

In a live-streaming context, results suggested that the tool has promise, in particular for casual streaming audiences, and perhaps also for female audiences. I also found that All the Feels seemed to increase perceived connection between the spectator and the streamer, while offering an engaging and enjoyable experience for spectators. The results of this research support the premise that biometric data could be used as an additional layer of information to spark conversation and enhance streamer and spectator experience. All the Feels is a promising addition to gameplay streams, to be proven out in more in-depth, extended field testing with the tool in various streamers' environments. I am currently conducting more studies using the tool in a horror-game genre, as well as creating more customization options and features, such as different themes of emojis, for the streamer to add to the overlay.

Next, I used the tool to gauge how participants felt about a game, compared with other methods to evaluate player experience. The study included multiple methods for evaluating players' emotional responses to a game, and when used together, they capture a larger picture of emotional variance among players. The results showed differences in individual expressivity. For example, for some users who were not very expressive verbally or nonverbally, biometrics were nonetheless able to pick up peak emotional responses. The tool was unobtrusive to the player, while still capturing a large amount of biometric data throughout play, which supports the early design goals for the tool [29]. Because the findings of this research support using a mix of methods when evaluating player experience, for future work I hope to create an easily accessible 'kit' for game researchers so they can use the SEI and ATF in studies of their own.

This thesis shows the value of this tool within the larger context of games user research in addition to demonstrating potential applications in contexts outside of games. I see it as a tool to evaluate how people respond emotionally for researchers in a wide range of fields.

Bibliography

[1] Twitch.

[2] Twitch Plays Pokemon.

[3] Madeline Alsmeyer, Rosemary Luckin, and Judith Good. Developing a Novel Interface for Capturing Self Reports of Affect. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '08, pages 2883–2888, New York, NY, USA, 2008. ACM.

[4] Kirsten Boehner, Rogrio DePaula, Paul Dourish, and Phoebe Sengers. How Emotion is Made and Measured. *Int. J. Hum.-Comput. Stud.*, 65(4):275–291, 4 2007.

[5] Florian Brühlmann and Gian-Marco Schmid. How to Measure the Game Experience?: Analysis of the Factor Structure of Two Questionnaires. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, pages 1181–1186, New York, NY, USA, 2015. ACM.

[6] Frank Coffield, David Moseley, Elaine Hall, and Kathryn Ecclestone. *Learning*

styles and pedagogy in post-16 learning: a systematic and critical review. Learning & Skills Research Centre, London, England, 2004.

[7] Alena Denisova, A Imran Nordin, and Paul Cairns. The Convergence of Player Experience Questionnaires. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '16, pages 33–37, New York, NY, USA, 2016. ACM.

[8] Ansgar E Depping and Regan L Mandryk. Why is This Happening to Me?: How Player Attribution Can Broaden Our Understanding of Player Experience. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 1040–1052, New York, NY, USA, 2017. ACM.

[9] J. Dewey. *Art as experience.* Penguin, 1934.

[10] Albert H. Huang, David C. Yen, and Xiaoni Zhang. Exploring the potential effects of emoticons. *Information & Management*, 45(7):466–473, 2008.

[11] Katherine Isbister. *How Games Move Us: Emotion by Design.* MIT Press, 2016.

[12] Katherine Isbister, Kristina Höök, Michael Sharp, and Jarmo Laaksolahti. The Sensual Evaluation Instrument: Developing an Affective Evaluation Tool. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 1163–1172, New York, NY, USA, 2006. ACM.

[13] Charlene Jennett, Anna L Cox, Paul Cairns, Samira Dhoparee, Andrew Epps, Tim

Tijs, and Alison Walton. Measuring and Defining the Experience of Immersion in Games. *Int. J. Hum.-Comput. Stud.*, 66(9):641–661, 9 2008.

[14] Madison Klarkowski, Daniel Johnson, Peta Wyeth, Cody Phillips, and Simon Smith. Psychophysiology of Challenge in Play: EDA and Self-Reported Arousal. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, pages 1930–1936, New York, NY, USA, 2016. ACM.

[15] Patrick Klepek. Who Invented Let's Play Videos?, 2015.

[16] M.L. Knapp and J.A. Hall. *Nonverbal communication in human interaction*. Wadsworth Thomson Learning, Australia, 2002.

[17] Rachel Lee Knickmeyer and Michael Mateas. Preliminary Evaluation of the Interactive Drama Facade. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '05, pages 1549–1552, New York, NY, USA, 2005. ACM.

[18] Jarmo Laaksolahti, Katherine Isbister, and Kristina Höök. Using the Sensual Evaluation Instrument. *Digital Creativity*, 20(3):165–175, 2009.

[19] B. Laurel. *Computers as theatre*. 1993.

[20] Fabien Lotte. Brain-Computer Interfaces for 3D Games: Hype or Hope? In *Foundations of Digital Games (FDG'2011)*, pages 325–327, Bordeaux, France.

[21] Regan L. Mandryk and Kori M. Inkpen. Physiological Indicators for the Evaluation of Co-located Collaborative Play. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*, CSCW '04, pages 102–111, New York, NY, USA, 2004. ACM.

[22] Regan L. Mandryk, Lennart E. Nacke, and Regan L. Mandryk. Biometrics in Gaming and Entertainment Technologies. In *Biometrics in a Data Driven World: Trends, Technologies, and Challenges*, chapter 6, pages 191–224. CRC Press, 2016.

[23] Pejman Mirza-Babaei, Lennart Nacke, Geraldine Fitzpatrick, Gareth White, Graham McAllister, and Nick Collins. Biometric Storyboards: Visualising Game User Research Data. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '12, pages 2315–2320, New York, NY, USA, 2012. ACM.

[24] J.H. Murray. *Hamlet on the holodeck: the future of narrative in cyberspace.* MIT Press, 1997.

[25] Lennart E. Nacke, Michael Kalyn, C Lough, and Regan L. Mandryk. Biofeedback game design: using direct and indirect physiological control to enhance game interaction. *Proceedings of the SIGCHI ...*, pages 103–112, 2011.

[26] Maartje M C G Polman, Licia Calvi, and Dirk Janssen. Biometric Design for Casual Games: A Case Study on Measuring Facial Responses in Casual Games. In *Proceedings of the 8th International Conference on Advances in Computer En-*

tertainment Technology, ACE '11, pages 67:1–67:2, New York, NY, USA, 2011. ACM.

[27] Reddit. what happened to the heart monitors in tournaments?, 2012.

[28] Stuart Reeves, Sarah Martindale, Paul Tennent, Steve Benford, Joe Marshall, and Brendan Walker. The Challenges of Using Biodata in Promotional Filmmaking. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 22(3):11, 2015.

[29] Raquel Robinson, Katherine Isbister, and Zachary Rubin. All the Feels : Introducing Biometric Data to Online Gameplay Streams. In *CHI PLAY '16 Extended Abstracts*, pages 261–267, 2016.

[30] Raquel Robinson, John Murray, and Katherine Isbister. "You're Giving Me Mixed Signals!": A Comparative Analysis of Methods that Capture Players' Emotional Response to Games. In *Forthcoming CHI '18 Extended Abstracts*, Montreal, QC, Canada, 2018. ACM.

[31] Raquel Robinson, Zachary Rubin, Elena Mrquez Segura, and Katherine Isbister. All the Feels: Designing a Tool That Reveals Streamers' Biometrics to Spectators. In *Proceedings of the 12th International Conference on the Foundations of Digital Games*, FDG '17, pages 36:1–36:6, New York, NY, USA, 2017. ACM.

[32] Katie Salen and Eric Zimmerman. *Rules of Play: Game Design Fundamentals*. The MIT Press, 2003.

[33] Holger Schnädelbach, Stefan Rennick Egglestone, Stuart Reeves, Steve Benford, Brendan Walker, and Michael Wright. Performing thrill: designing telemetry systems and spectator interfaces for amusement rides. In *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*, number February 2017, page 1167, 2008.

[34] N Sadat Shami, Jeffrey T Hancock, Christian Peter, Michael Muller, and Regan Mandryk. Measuring Affect in Hci: Going Beyond the Individual. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '08, pages 3901–3904, New York, NY, USA, 2008. ACM.

[35] Jaakko Stenros, Janne Paavilainen, and Frans Mäyrä. The many faces of sociability and social play in games. In *Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era on - MindTrek '09*, page 82, 2009.

[36] Chek Tien Tan, Sander Bakkes, and Yusuf Pisan. Inferring Player Experiences Using Facial Expressions Analysis. In *Proceedings of the 2014 Conference on Interactive Entertainment*, IE2014, pages 7:1–7:8, New York, NY, USA, 2014. ACM.

[37] Paul Tennent, Stuart Reeves, Steve Benford, Brendan Walker, Joe Marshall, Patrick Brundell, Rupert Meese, and Paul Harter. The machine in the ghost. In *Proceedings of the 2012 ACM annual conference extended abstracts on Human Factors in Computing Systems Extended Abstracts - CHI EA '12*, page 91, 2012.

[38] Arnold P O S Vermeeren, Effie Lai-Chong Law, Virpi Roto, Marianna Obrist,

Jettie Hoonhout, and Kaisa Väänänen-Vainio-Mattila. User Experience Evaluation Methods: Current State and Development Needs. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, NordiCHI '10, pages 521–530, New York, NY, USA, 2010. ACM.

[39] Chen Wang, Erik N. Geelhoed, Phil P. Stenton, and Pablo Cesar. Sensing a live audience. In *Chi '14*, pages 1909–1912, 2014.