Aalborg University Copenhagen

Aalborg University Copenhagen

Frederikskaj 12,

DK-2450 Copenhagen SV

Semester Coordinator:
Stefania Serafin

Secretary:
Lisbeth Nykjær

Semester:
10th Semester Medialogy

Title:
HeartBee: A Player Experience Evaluation of a Mobile HRV Biofeedback Game

Project Period:
Summer 2018 (May 31st – August 31st)

Semester Theme:
Master Thesis

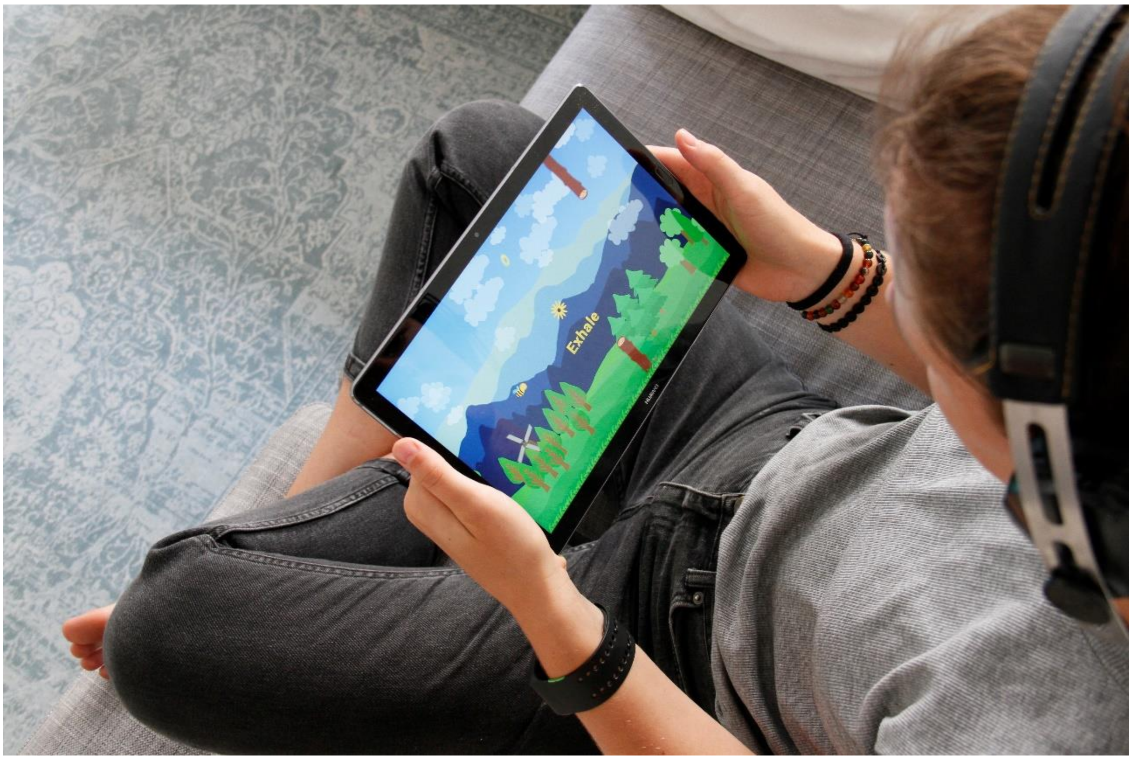
Supervisor(s):
Henrik Schønau-Fog

Members:
Rasmus Kongsmar Brejl
20132736

Abstract:
Heart rate variability (HRV) biofeedback training is an established procedure for reducing stress and bringing balance to the autonomous nervous system by performing slow breathing exercises. This procedure, however, can become monotonous and boring. HRV biofeedback games are a category of games that try to make the experience more exciting and engaging. However, there is a lack of documentation on whether they improve the experience.
This thesis sought out to test if HeartBee, an HRV biofeedback game, could improve engagement in biofeedback training. Engagement was defined as a combination of flow, continuation desire, and time spent playing.
The game was evaluated on 31 participants utilizing a counterbalanced repeated-measures experiment. The game was compared to a non-game HRV biofeedback application.
Flow scores were measured using the Short Flow State Scale (SFSS). Continuation desire was measured using the basic game ESQ. Participants were free to end the experiences at any time they wanted, and as such, time spent was measured from when they started their experience until they ended it.
The results confirm the potential of using games for engaging people in biofeedback training. Participants reported higher levels of flow and continuation desire during HeartBee and they played it significantly longer than they used the non-game application. The results of this thesis show the potential of using serious games to engage people in healthy activities.

HEARTBEE

A PLAYER EXPERIENCE EVALUATION OF A MOBILE HRV BIOFEEDBACK GAME



A Master's thesis by Rasmus Kongsmar Brejl

Aalborg University Copenhagen

August 2018

ABSTRACT

Heart rate variability (HRV) biofeedback training is an established procedure for reducing stress and bringing balance to the autonomous nervous system by performing slow breathing exercises. This procedure, however, can become monotonous and boring. HRV biofeedback games are a category of games that try to make the experience more exciting and engaging. However, there is a lack of documentation on whether they improve the experience.

This thesis sought out to test if HeartBee, an HRV biofeedback game, could improve engagement in biofeedback training. Engagement was defined as a combination of flow, continuation desire, and time spent playing.

The game was evaluated on 31 participants utilizing a counterbalanced repeated-measures experiment. The game was compared to a non-game HRV biofeedback application.

Flow scores were measured using the Short Flow State Scale (SFSS). Continuation desire was measured using the basic game ESQ. Participants were free to end the experiences at any time they wanted, and as such, time spent was measured from when they started their experience until they ended it.

The results confirm the potential of using games for engaging people in biofeedback training. Participants reported higher levels of flow and continuation desire during HeartBee and they played it significantly longer than they used the non-game application. The results of this thesis show the potential of using serious games to engage people in healthy activities.

Preface

This project is written as a master's thesis in Medialogy – Games at Aalborg University Copenhagen.

The motivation of this thesis originally stemmed from my passion for developing serious games. A previous version of HeartBee was developed with the purpose of improving players' HRV. The results, from an HRV perspective, were promising but I realized that most serious games focused only on the clinical effects of the games rather than the player experience of the games. So how do we measure the player experience in such games? It is my passion to find out and to verify that games can indeed be engaging while conveying serious topics.

My previous work, "HeartBee: Development and Evaluation of a Personalized HRV Biofeedback Game" is an unpublished bachelor project. It will be referenced throughout this thesis and have therefore been added to the digital appendix.

This thesis is accompanied by the mobile game HeartBee. The game is developed for Android and utilizes an Empatica E4 wristband for playing. HeartBee can also be found in the digital appendix.

The thesis is showcased through the following A/V production:

https://www.youtube.com/watch?v=ibZQVRZ_q3A&feature=youtu.be

I would like to thank Daniel H. Ditlevsen and Alexander K. Risvang for helping me film, edit, and act in the A/V production that accompanies this thesis.

Thank you to the SMILE lab and Augmented Cognition Lab at AAU CPH for letting me borrow equipment for developing and testing HeartBee.

Thank you to Kayla Friedman and Malcolm Morgan of the Centre for Sustainable Development, University of Cambridge, UK for producing the Microsoft Word thesis template used to produce this document.

CONTENTS

LIST OF TABLES

List of Figures

List of Abbreviations and Acronyms

ANS	Autonomous Nervous System
BLE	Bluetooth Low Energy
CD	Continuation Desire
ECG	Electrocardiography
EDA	Electrodermal Activity
EEG	Electroencephalography
EMG	Electromyography
ESM	Experience sampling method
ESQ	Engagement Sample Questionnaire
FSS	Flow State Scale
GengQ	Game Engagement Questionnaire
GexpQ	Game Experience Questionnaire
HF	High Frequency
HR	Heart Rate
HRCP	HeartRate+ Coherence Pro
HRV	Heart Rate Variability
IBI	Interbeat Interval
mHealth	Mobile Health
PPG	Photoplethysmography
RMSSD	Root Mean Square of Successive Differences
RSA	Respiratory Sinus Arrhythmia
SDNN	Standard Deviation of the normal beat to beat interval
SFSS	Short Flow State Scale
UI	User Interface

1 INTRODUCTION

This thesis is centred around “HeartBee” which is a biofeedback game I developed in a previous project.

HeartBee is a Mobile health (mHealth) application. mHealth applications are mobile applications that are used for either treating diseases or as prevention. This is done e.g. through promoting physical and/or dietary lifestyle changes, or through biofeedback. The mHealth market is a rapidly growing one, with 325.000 mHealth applications available in 2017 from 84.000 publishers (Pohl, 2017). Only a fraction of these publishers, however, reach a broad target group and has a significant revenue (Research 2 Guidance, 2018).

An analysis of the usage of mHealth applications found that people who were younger (18-44 years old), had a higher education (college or higher), had higher income, and reported better general health were more likely to use mHealth applications (Carroll, et al., 2017).

Lister et al. (2014) analysed 132 health and fitness applications utilizing gamification. They found that gamification was only used to some extent in most applications, but that game elements were correlated with app popularity. Their findings suggest that health-app developers should create in-depth gaming experiences rather than only using gamification when convenient (Lister, West, Cannon, Sax, & Brodegard, 2014).

The video game industry is a rapidly growing industry, expected to reach a global revenue of $137.9 Billion in 2018 according to a games market report by Newzoo (2018). In the same report, it is estimated that there are 2.3B active gamers in the world. Of this revenue, more than 50% come from mobile games, making it the primary game platform.

With games dominating today’s entertainment industry, it is inevitable that some games succeed while others do not. The success of some games compared to others can be attributed to many variables such as budget, marketing, and timing. A variable that has sparked interest in scholars, however, is player experience.

Player experience is a term that refers to how the player experiences playing the game and it is closely related to player psychology (Nacke & Drachen, 2011). In general terms, game developers want their users to have a good experience playing their game. But what is a good experience, and how do we evaluate it?

Specifically related to mHealth applications as well as games for health, studies are primarily focused on evaluating the health benefits of the application or game, and there is an abundance of studies in evaluating the player experience of such applications or games.

My mHealth application HeartBee proved effective at increasing short term resting heart rate variability (HRV) in players after only 10 minutes of playing (Brejl, 2016). However, HeartBee utilized sensor technology that only allowed it to be played on a computer, largely limiting the mobility of the application.

An assumption is that mHealth applications need to engage their users to be effective. Even the best technical mHealth application has no effect if its users will not use it. It is my basic assumption that by designing and applying game elements properly, developers can ensure engagement and user retention.

With mobile games dominating the game industry, the purpose of this thesis is thus two-fold:

First, HeartBee should see further design iterations and be developed into a mobile application to reach a broader target group.

Secondly, the player experience of HeartBee should be measured through standard player experience measurements and compared to player experience measurements of an mHealth application. The two objectives of this thesis can thus be summarized in the following initial problem statement:

> *How can a mobile HRV biofeedback game increase the user experience during biofeedback training?*

I will in this thesis research how player experience can be defined and measured. The definition of player experience will then be used to propel the design of the biofeedback with the purpose of optimizing the experience. First, however, I will cover the background of HRV biofeedback training.

2 BACKGROUND

This section aims to cover the academic background through literature reviews. The first section is dedicated to cover the background of heart rate variability with the purpose of providing the reader with an understanding of the underlying psychophysiological constructs in biofeedback training.

The second section deals with examining the literature on what makes a good player experience. The purpose of this section is to define what kind of experience applies to HeartBee.

The third section deals with exploring how the defined player experiences can be measured with a focus of finding methods that can be used to evaluate HeartBee.

2.1 Heart rate variability

HeartBee is a biofeedback game controlled by the heart rate. To understand the controls and design, this section aims to cover the background on heart rate and heart rate variability, why it is important to improve heart rate variability, and how this is achieved through biofeedback. This chapter is largely based on my previous project (Brejl, 2016).

2.1.1 Autonomic nervous system and heart rate variability

Heart rate variability (HRV) refers to the variation in interbeat intervals (IBI) of the heart rate (HR).

The heart is regulated by two branches of the autonomic nervous system (ANS) - the parasympathetic nervous system and the sympathetic nervous system (Shaffer, McCraty, & Zerr, 2014). The parasympathetic nervous system is responsible for the rest and digest response whereas the sympathetic nervous system is responsible for the fight or flight response. HR is modulated by the relative activity of the two branches - a relative increase in the sympathetic nervous system increases HR and lowers HRV. A relative increase in the parasympathetic nervous system decreases HR and increases HRV (Lane, et al., 2009).

The parasympathetic nervous system is dominated by the vagus nerve. Parasympathetic activity is therefore also referred to as vagal activity, vagal tone, vagal functioning, or cardiac vagal control (Laborde, Mosley, & Thayer, 2017). The vagus nerve is stimulated through respiration where inspiration increases HR and expiration decreases HR (Lane, et al., 2009). The modulation of the HR through respiration is also known as respiratory sinus arrhythmia (RSA).

2.1.2 Measuring HRV

Sympathetic and vagal activity are generally evaluated through measuring HRV.

HR is usually measured either through electrocardiography (ECG) or through photoplethysmography (PPG) (Laborde, Mosley, & Thayer, 2017). ECG obtains direct data through measuring the electrical stimulus of the heart. PPG obtains data through shining a light onto the skin and measuring the reflected light. The reflected light depicts blood volume which is dependent on the heartbeat. Of the two measurements, ECG is considered the ground truth and is preferred in clinical studies of HRV.

HRV is usually analysed using either frequency domain measurements or time domain measurements (Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology., 1996). Frequency domain measurements divide a HR measurement into four frequency components - the high frequency (HF) (0.15 Hz - 0.4 Hz), the low frequency (LF) (0.04 Hz - 0.15 Hz), the very-low frequency (VLF) (0.0033 Hz - 0.04 Hz), and the ultra-low frequency (ULF) component (<0.0033 Hz).

The HF component reflects vagal activity whereas the LF component mainly reflects sympathetic activity. The LF component, however, can also reflect vagal activity during slow breathing where the respiratory frequency moves into the LF component. The VLF and ULF components do not reflect vagal activity and are therefore not of interest for this project.

The most commonly reported time domain measurements include the standard deviation of the normal beat to beat interval (SDNN) and the root mean square of successive differences (RMSSD). SDNN is believed to reflect all frequency components but is mainly accounted for by RSA during short term recordings. RMSSD reflects beat to beat variance and therefore strongly reflects the HF component and vagal activity (Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology., 1996).

Vagal activity is a strong indicator of overall health and is linked to a person’s psychophysiology. Psychophysiology is a term that describes the physiological bases of psychological processes.

Vagal activity decreases with age (Meersman & Stein, 2007) but reduced vagal activity has also been associated with stress, panic, anxiety, and worry (Shaffer, McCraty, & Zerr, 2014), reduced self-regulatory capacity and cognitive function (McCraty & Shaffer, Heart Rate Variability: New Perspectives on Physiological Mechanisms, Assessment of Self-regulatory Capacity, and Health risk, 2015), an increased risk for morbidity and mortality (F.Thayer & D.Lane, 2007), as well as general poor health outcomes like diabetes, obesity, hypertension and other cardiovascular diseases (Chambers & Allen, 2007). On the other hand, higher levels of vagal activity are associated with attention (Chambers & Allen, 2007) as well as higher cognitive function (Thayer, Hansen, Saus-Rose, & Johnsen, 2009).

When measuring HRV it is important to assess all the extrinsic factors that can affect HR. Laborde et al. (2017) state that researchers should consider two types of variables - stable and transient variables. Stable variables that should be considered are:

- Age and gender
- Smoking
- Alcohol habits
- Weight, height and hip-to-waist ratio
- Cardioactive medicine
- Oral contraceptive intake for females

Transient variables that should be considered are:

- Normal sleeping routine
- No intense physical training the day before the experiment
- No meal in the last 2 h before the experiment
- No coffee or caffeinated drinks in the last 2 h before the experiment
- No alcohol in the last 24 h before the experiment.

All the listed variables affect HRV and should therefore be accounted for when comparing HRV in studies (Laborde, Mosley, & Thayer, 2017). In addition to those variables, the Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology (1996) state that HRV should only be compared between sessions of equal length. As such, there are a lot of challenges within measuring and comparing HRV. The focus of this thesis, however, is not on comparing HRV measurements and the variables and guidelines do therefore not need to be controlled for (Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology., 1996).

2.1.3 HRV Biofeedback training and psychophysiological coherence

Biofeedback is a procedure in which individuals receive information about their physiology. During biofeedback training, they can try to get their physiology under control through self-regulation. Biofeedback training is normally done using a screen (e.g. a computer screen) for feedback. Over time, they can learn to self-regulate without monitoring the process with a screen (Frank, Khorshid, Kiffer, Moravec, & McKee, 2010).

In relation to HRV and vagal activity, biofeedback training aims at bringing control over the HR through slow deep breathing.

HRV biofeedback training, also known as RSA biofeedback or resonance frequency feedback, is a procedure in which individuals are asked to perform slow breathing exercises while following their HR on a screen (Lehrer, Vaschillo, & Vaschillo, 2000). The task of the individuals is to maximise their RSA amplitude while making their HR follow a sinusoidal curve. Making your HR follow this curve can be achieved by almost anyone in less than a minute, regardless of prior training (Lehrer & Gevirtz, 2014).

Lehrer & Gevirtz argue that optimal results, i.e. maximum RSA amplitude, is achieved when the HR and respiration synchronize such that the HR and respiration are in phase. This in-phase relationship is found in individuals' resonance frequency - an internal frequency in which systems like blood pressure oscillate. The resonance frequency is generally found at 0.1 Hz which amounts to 6 breaths per minute (Lehrer & Gevirtz, 2014) but is unique for each individual and may vary from 4.5 breaths per minute to 7 breaths per minute (Steffen, Austin, DeBarros, & Brown, 2017).

Psychophysiological coherence is a theory that expands upon HRV biofeedback training. McCraty & Shaffer (2015) describe physiological coherence as a synchronized relationship between internal systems, such as HR, BP, and respiration rate. This is indeed very similar to the resonance frequency breathing as described earlier. However, psychophysiological coherence extends this theory to include emotions. McCraty et al. have found a correlation between physiological coherence and positive emotions such as appreciation and compassion. This contrasts with negative emotions like stress, anxiety, and fear, which they have correlated to incoherent physiology (McCraty, Atkinson, Tiller, Rein, & Watkins, 1995; McCraty & Childre, 2010). The findings of McCraty et al. even suggest that positive emotions automatically form sine-wave like pattern in the HR without a conscious change of breathing pattern. In general, the findings of McCraty et al. indicate that being in a coherent state correlates with a sense of wellbeing and that it improves cognitive, social, and physical performance (McCraty & Shaffer, Heart Rate Variability: New Perspectives on Physiological Mechanisms, Assessment of Self-regulatory Capacity, and Health risk, 2015).

Optimal effects on HRV are achieved during psychophysiological coherence in which the physiology is controlled through resonance frequency breathing while the individual focuses on positive emotions. The effects of emotions on the HR are shown in Figure 1 as both beat to beat variance and as the power spectral density.

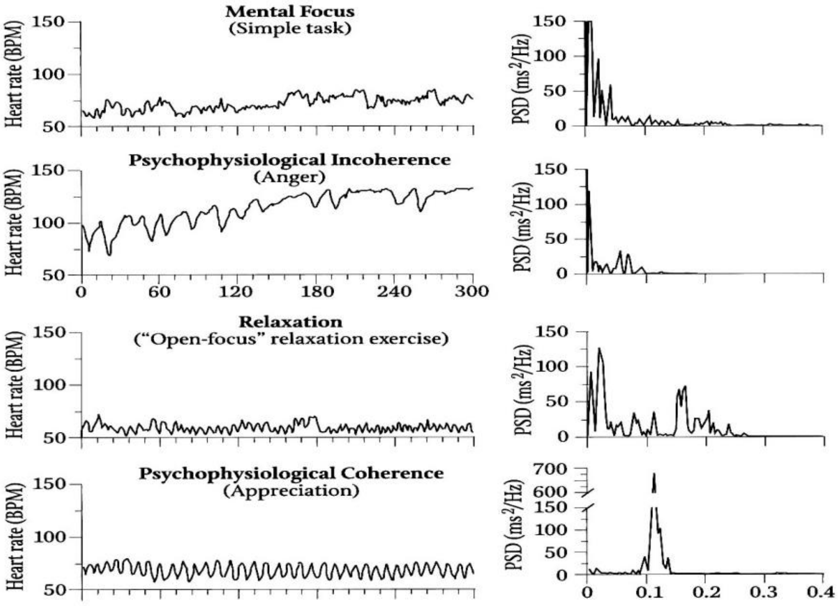


Figure 1 - An illustration of how emotions affect the HR (McCraty & Childre, 2010).

2.1.4 Effects of HRV biofeedback training

HRV biofeedback training and psychophysiological coherence training have both seen success in clinical studies:

A study of the effects of 10 minutes of HRV biofeedback training on cognitive performance during laboratory induced stress through a modified Stroop task found that the group who received biofeedback training showed faster reaction time, less mistakes, and a stronger consistency in their responses compared to a control group. In addition to that, the biofeedback group was more relaxed, less anxious, and less sleepy (Prinsloo, et al., 2010).

A study of combat veterans suffering from PTSD found that weekly psychophysiological coherence training with a professional for four weeks improved their post-intervention HRV significantly (Ginsberg, Berry, & Powell, 2010).

A study on the effect of HRV biofeedback training on performance psychology of basketball players found that 20 minutes of daily training for 10 consecutive days led to increased HRV, reduced anxiety traits, and improvements in basketball related tasks compared to a control group. Their findings were even consistent during a follow-up measure one month later (Paul & Garg, 2012).

The findings of my previous study suggested that playing a HRV biofeedback game for 10 minutes increased short term resting HRV significantly compared to a control group (Brejl, 2016).

2.1.4.1 HRV Biofeedback games

With the clinical success of HRV biofeedback training, it is no surprise that there has been an emergence in products for biofeedback training at home.

The HeartMath Institute (2018) is perhaps the leader in products for HRV biofeedback training at home with a series of products for both computer, android, and iOS. Most of their products utilize a non-invasive earlobe or finger pulse sensor.

The benefits of HRV biofeedback training is like other exercises - more time and more sessions lead to better results (Wollmann, et al., 2016). An assumption that could be made is that monotonous breathing at the same frequency over an extended period can lead to boredom, which makes it hard to keep individuals engaged in the task, despite the benefits of the training.

One way to engage people, however, is through gamification of the training or even turning the training into a serious game. Somatic Vision and COMM Studios are both companies that specialize in developing HRV biofeedback games. COMM Studios focuses on more relaxed applications with the HR being the only interaction whereas Somatic Vision provides more interactive experiences through games like “Tropical Heat” where the HR interaction becomes secondary to more standard game interactions (TheBiofeedback; Somatic Vision, 2018).

For both cases, like with many serious games, the focus of evaluation is on clinical studies that prove the games’ effect. Oftentimes, it is only assumed that, because it is a game, it engages its users by default and player experience testing receives little attention.

The study by Wollmann et al. (2016) is the only study I have been able to identify that aims to evaluate the user experience of a biofeedback game. In their study, they follow a user-centred design approach for developing their game. Their findings suggest that users prefer games with more elements and interaction, but that the clinical benefits of HRV biofeedback are larger when the interaction focuses only on biofeedback training. Their sample, however, was limited to 11 participants ranging in age from 40 to 67 years old (Wollmann, et al., 2016).

To my knowledge, no study has been made comparing the user experience of a non-game HRV biofeedback application (limited to a data-focused visualization of the HR) to the user experience of a HRV biofeedback game. Doing so can not only shed light on the power of serious games for engaging individuals in improving their own health but can also validate the results of Wollmann et al. (2016) and help propel the development of good HRV biofeedback games.

2.2 Defining player experience

With the purpose of evaluating user experience in a HRV biofeedback system, it is important to first define how a user experience is described by different theories. With the focus of this thesis being on games, user experience will henceforth be referred to as player experience.

Most player experience theories focus on defining certain mental states of a player that can be measured through subjective measures, objective measures, or a combination of both. Terms for describing what engages a player in a game vary widely in literature and include but are not limited to immersion (Ermi & Mäyrä, 2005; Jennett, et al., 2008), presence (Lombard & Ditton, 1997), fun (Koster, 2004), enjoyment (IJsselsteijn, et al., 2008), flow (Csikszentmihalyi, 1990), and continuation desire (Schoenau-Fog, 2011b). In this section, I review three distinct theories of an engaging player experience, namely immersion, continuation desire, and flow.

2.2.1 Immersion

Immersion is a widely used term when talking about an engaging player experience. An early definition of immersion by Janet Murray (1997) describes immersion as "*a metaphorical term derived from the physical experience of being submerged in water*" (Murray, 1997). This means that immersion is the sensation of being completely absorbed by the game and being surrounded by a different reality. This definition of immersion was enhanced by McMahan (2003) to encompass the relationship between immersion and engagement.

In her work, McMahan mentions two different conventions of immersion. One is that the player is caught up in the world of the game's story. She refers to this as the diegetic level.

The second is the player's love of the game and the strategy that goes into it. She refers to this as the non-diegetic level.

She explains that photo- and audio-realism is not a necessity for experiencing immersion. Instead, she describes immersion as the sense of the virtual world being real and complete.

Immersion, according to McMahan, is defined by the previously described diegetic level.

She refers to the non-diegetic level as engagement. She describes engagement as gaining points, devising a strategy, and showing off to other players both during and after gameplay. In some cases, she refers to engagement as *deep play*. This term is referred to as being so engaged in a game that a player reaches a level of near-obsessiveness. She goes on to describe that deep play, in the sense of video games, is a measure of a player's level of engagement (McMahan, 2003).

Brown & Cairns (2004) define immersion in games as a three-step process consisting of engagement, engrossment, and total immersion.

They refer to engagement as the lowest level of involvement with the game which must occur before any of the other levels can happen. They identify two barriers to enter engagement - one being access. Access is referred to as the player preferences (does the player like the style of game) and controls of the game. In relation to controls, it is important that the player can become an expert at the main controls.

The second barrier is the investment of the player. The player needs to invest time, effort, and attention into the game. Once both barriers have been lowered, the player can become engaged with the game.

The second level of immersion, engrossment, is defined as the stage at which the player's emotions are affected by the game. The barrier to engrossment is the game construction, e.g. the visuals, sounds, tasks, and plot of the game.

During engrossment, Brown & Cairns state that "*the game becomes the most important part of the gamer's attention and their emotions are directly affected by the game*" (p. 1299). In addition to this, the gamers become less aware of their surroundings and their self.

They define the third and final level of immersion, total immersion, as presence. At the stage of total immersion, players described that the game was the only thing that impacted their thoughts and feelings. As such, they were detached from reality and "present" in the game. They define the barriers to total immersion as empathy and atmosphere. Empathy is described as an emotional attachment to the characters of the game. Atmosphere is an expansion of game construction where the features of the game need to be relevant to the actions and locations of the game characters.

Finally, they state that attention is an important part of total immersion and they divide it into visual, auditory, and mental attention. They state that immersion is correlated with the amount of attention and effort invested into the game (Brown & Cairns, 2004).

Ermi & Mäyrä (2005) define immersion as a three-dimensional construct - sensory immersion, challenge-based immersion, and imaginative immersion.

They define sensory immersion as immersion coming from the audio-visual landscape of the game. They state the stimuli can easily overpower the sensory information of the real world which makes the player capable of focusing all attention on the game.

They define challenge-based immersion as the immersion that arises from solving tasks that have a satisfying balance between challenge and ability. This dimension of immersion is closely related to flow.

They define imaginative immersion as the immersion that arises from becoming absorbed with the stories and the world or beginning to feel for or identify with characters in the game (Ermi & Mäyrä, 2005). Their definition of imaginative immersion is closely related to the definition of total immersion by Brown & Cairns (2004) and the definition of immersion by McMahan (2003).

Through their investigations of immersion in video games, Jennett et al. (2008) argue that immersion differs from related terms like presence, flow, and cognitive absorption. They argue that, while flow describes an optimal experience, immersion is only concerned with "*the specific, psychological experience of engaging with a computer game*" (p. 643). They mention that people do not always play video games with the purpose of being immersed, but that immersion rather just happens as they get drawn into the game. They argue, however, that immersion is key to a good gaming experience (Jennett, et al., 2008).

Immersion was thus presented from different theories. McMahan (2003) defined immersion as being engaged with the narrative of the game. Brown & Cairns (2004) defined immersion as a three-step process from engagement to engrossment to total immersion. Their definition of total immersion was also related to the narrative. Ermi & Mäyrä (2005) defined immersion as a three-dimensional construct: sensory-based immersion, challenge-based immersion, and imaginative immersion. Sensory-based immersion was related to the audio-visuals of the game, challenge-based immersion was related to the balance between skills and challenge, and imaginative immersion was related to the narrative of the game. Finally, Jennett et al. (2008) defined immersion as the psychological experience of engaging with a computer game.

HeartBee is not a narrative based game as it lacks story and character development. The definitions of immersion that are interesting for this thesis are therefore the sensory-based and challenge-based immersion by Ermi & Mäyrä (2005) as well as the engagement definition by Jennett et al. (2008).

2.2.2 Engagement as Continuation Desire

Player engagement can be understood as a player's level of wanting to continue playing while in a game, or as the player's want to come back to the game. Schoenau-Fog (2011b) defines this desire to continue playing a game as continuation desire.

In his work, Schoenau-Fog distinguishes between motivation and engagement, stating that motivation is the reason why people start playing and get lured in, whereas engagement is concerned with what makes the players "hooked" and want to continue playing the game. The two are distinguished as the motivation to start playing is not necessarily related to being engaged in the game but could also be related to general boredom (i.e. having nothing better to do). Player engagement can be described as a combination of the two - a player's individual motivation to begin playing as well as the desire to keep playing. Schoenau-Fog states that player engagement, in this sense, is related to emotions like enjoyment, fun, and satisfaction, as well as experiences such as being immersed, in presence, or in flow.

With a focus of identifying engagement and disengagement triggers, Schoenau-Fog used a grounded theory approach to construct the OA3 framework which consists of the categories Objectives, Activities, Accomplishments, and Affect. The OA3 categories are made up of 18 conceptual categories. Figure 2 depicts the O3 framework as well as the relationship between the categories.

The four categories of the OA3 framework are all intrinsically connected. Schoenau-Fog explains this through the player engagement process.

The player engagement process begins with the player's motivation to start playing. This motivation can either be intrinsic (e.g. wanting to meet new friends), extrinsic (e.g. suggested by friends), or a mix thereof.

When entering the game, the player is usually presented with extrinsic objectives by the game. An extrinsic objective could be to slay the dragon. Objectives are not limited to extrinsic objectives but can also be set up intrinsically by the player. A typical intrinsic objective could be the want to explore every corner of the game. Both extrinsic and intrinsic objectives are not static but can change dynamically throughout the game.

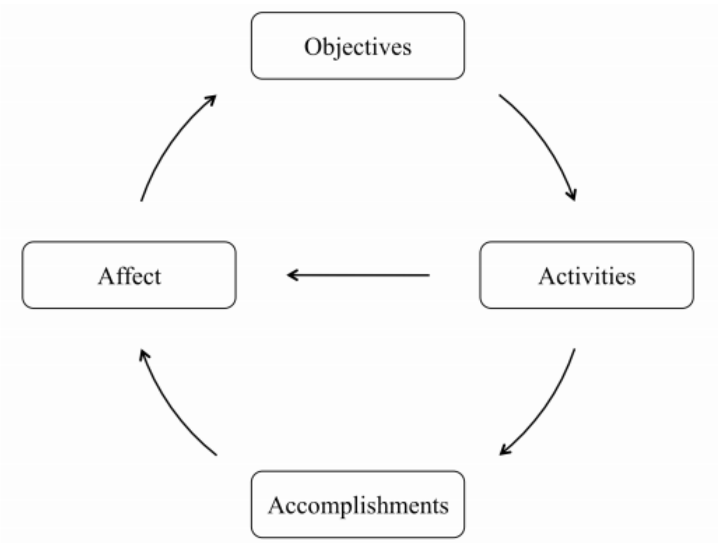


Figure 2 - The OA3 framework (Schoenau-Fog, 2011b).

With an objective in mind, the player can then start performing activities to accomplish the objectives. An example of this is to kill enemies (activity, labelled destruction in the framework) to gain experience and armour to progress the character (accomplishment). Other activities identified by Schoenau-Fog are: solving, sensing, interfacing, exploration, experimentation, creation, experiencing the story, experiencing the characters, and socializing. Other accomplishments identified are: achievements and completion.

Performing activities does not always lead to accomplishments but might lead to an affect instead. An example of this is the player being absorbed in killing enemies regardless of whether the character progresses or not. Other affects identified are: positive affect and negative affect.

Finally, the affect is evaluated and a subjectively positive affect for the player can lead to a desire to continue playing (or a desire to play again later), which leads to a new cycle with new objectives, activities, and accomplishments.

An extensive rank-ordered list of the 16 conceptual categories of activity, accomplishment, and affect (ranked based on how many times they were identified in respondents in Schoenau-Fog's work) with brief explanations is provided in Table 1.

Table 1 A brief overview of the 16 conceptual categories of the OA3 model (objectives excluded).

Category	Description
Accomplishment	
Completion	Refers to completing something in the game or finishing it fully. Players can stay engaged to complete everything in the game. Too high challenge or too low a challenge might lead to disengagement.
Progression	Refers to the sense of progressing the character of the game through gearing up, levelling up, or improving abilities. Can also refer to advancements such as points or scores A lack of variation and progression can lead to disengagement.
Achievement	Refers to the player wanting to achieve something in the game.

	No achievements in the game can lead to disengagement.
Activity	
Experiencing the story	Refers to players being engaged through experiencing the story and the progression of storylines. Too much story, boring or bad story, or no story at all can lead to disengagement.
Socializing	Refers to sharing the experience with others. Communication and competition can help drive engagement. Too much competition, or unbalanced competition, can lead to disengagement.
Sensing	Refers to the audio-visual landscape of the game. Bad or boring audio-visuals can lead to disengagement.
Exploration	Refers to exploring the game world and finding new experiences in the game. Getting stuck, being forced to explore, or when it takes too long time to get somewhere can lead to disengagement.
Experiencing the Characters	Refers to experiencing how characters evolve and develop through the game. Not being satisfied with characters in the game can lead to disengagement.
Solving	Refers to solving intellectual challenges, puzzles, problems, strategies, or tactics within the game. Too easy, too hard, or too repetitive puzzles can lead to disengagement.
Experimentation	Refers to the possibility of modifying the game, the mechanics, or the characters of the game.
Interfacing	Refers to how the players interact with the game through

	controls. Especially alternative control schemes can drive engagement. Badly designed controls can lead to disengagement.
Destruction	Refers to killing of NPCs and opponents as well as destruction of objects and structures. A lack of variation in destructive activities can lead to repetitiveness which can lead to disengagement.
Creation	Refers to the possibilities of users creating content and new elements in the game.
Affect	
Positive	Refers to positive emotions experienced in games, e.g. enjoyment, fulfilment, satisfaction, excitement. Positive affect also refers to physiological reactions, e.g. relaxation or adrenaline rush.
Absorption	Refers to a sense of flow, immersion, or presence in the game. A lack of immersion in the game can lead to disengagement.
Negative	Refers to negative response to the game, e.g. uninteresting game, boring, frustrating. Most negative affect leads to disengagement, but some affects, like frustration from losing, can lead to engagement through wanting to win the next time.

Schoenau-Fog validated the player engagement process through focused coding on answers from a separate study of engagement in World of Warcraft players. The results showed no new categories of player engagement, which means that his findings can be extended to other game genres (Schoenau-Fog, 2011b).

Continuation desire was thus defined as a player's desire to continue playing the game because of being engaged. The OA3 framework defined 18 conceptual categories that could drive engagement in a game. The OA3 framework can be used to analyse a game and find out where, why, and how players get engaged (or disengaged) with the game. In relation to this thesis it can therefore be used to better the design of HeartBee. In addition to that, continuation desire does not deal with defining a complex psychological construct but instead deals with a basic definition of engagement, i.e. the desire to continue.

2.2.3 Flow

Psychological flow, also referred to as flow, is a concept that Mihaly Csikszentmihalyi came across in his studies in pursuit of finding what made people happy in their lives. Csikszentmihalyi describes a flow experience as an optimal experience, an experience "*so gratifying that people are willing to do it for its own sake, with little concern for what they will get out of it, even when it is difficult or dangerous*" (Csikszentmihalyi, Flow: The psychology of optimal experience, 1990, p. 71). The flow experience is therefore an experience that is intrinsically rewarding and that, according to his interviews, seems to shadow external rewards (Csikszentmihalyi, 1975).

The flow experience is also described as the experience of "being in the zone". This experience, or the zone, is in its simplest sense achieved when the challenge of a given task equals the skill level of an individual. Only then is the individual capable of "losing oneself" in the task at hand. In contrast, a challenge that is too hard leads to anxiety whereas a challenge too easy leads to boredom (Csikszentmihalyi, 1975). Figure 3 depicts the relationship between challenge and skill.

The experience of flow, therefore, is very desirable and should be sought after in any activity that an individual does. The research of Csikszentmihalyi put forth nine components of enjoyment that constitute a flow experience (Csikszentmihalyi, Flow: The psychology of optimal experience, 1990): *A challenging activity that requires skills, the merging of action and awareness, clear goals, clear feedback, concentration on the task at hand, the paradox of control, the loss of self-consciousness, the transformation of time, and the autotelic experience*.

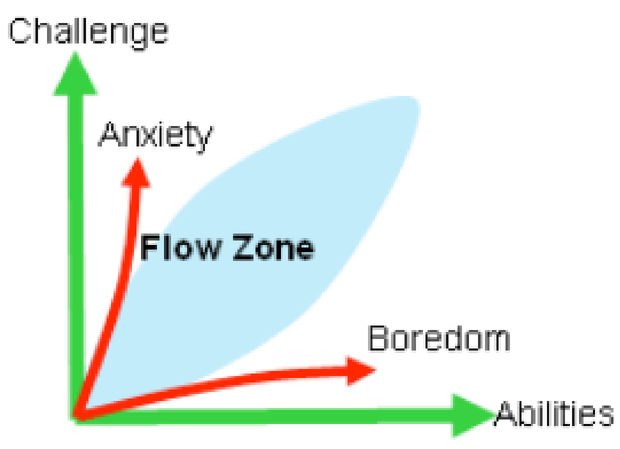


Figure 3 Flow as a function of challenge and abilities (skills) (Chen, 2007).

A challenging activity that requires skills is explained by the Flow channel model depicted in Figure 3. Without a challenge, the activity can lead to boredom and it becomes hard to lose oneself in it (Csikszentmihalyi, 1990).

The merging of action and awareness is described as the state where individuals are aware of their actions but not of the awareness itself (Csikszentmihalyi, 1975). This means that the individuals become so absorbed in their activity that they stop being aware of themselves outside of the activity.

Clear goals mean that the goal of the activity being performed should be clear to the individual. Csikszentmihalyi argues that setting a proper goal is very important, as trivial goals do not provide a sense of enjoyment (Csikszentmihalyi, Flow: The psychology of optimal experience, 1990).

Clear feedback means that the individual should always have clear direct feedback to evaluate a progression towards the goal. He states that for some activities, the goal and feedback may not be obvious, so the individual must be able to recognize and gauge feedback to enjoy the activity. It is impossible to reach flow without internal guidelines of how to achieve the goal through the feedback provided (Csikszentmihalyi, Flow: The psychology of optimal experience, 1990).

Concentration on the task at hand means that the activity should allow for a complete focus of attention on the task at hand, leaving no room for the mind wandering off (Csikszentmihalyi, 1990).

The paradox of control refers to the fact that a flow experience often involves a sense of control over the activity. Csikszentmihalyi adds that, more than a sense of control, the paradox of control is about not worrying about losing control. He adds that the paradox of control is more related to the possibility of control rather than the actuality of control (Csikszentmihalyi, Flow: The psychology of optimal experience, 1990).

The loss of self-consciousness is related to the second component: A flow experience so engrossing that there is no attention left to worry about the outside world also leads to the loss of self-consciousness. In a flow experience, individuals do not worry about how they do things, who they are, or how they are perceived from the outside world, but things instead become automatic with full attention on the task at hand (Csikszentmihalyi, Flow: The psychology of optimal experience, 1990).

The transformation of time is described as the feeling of a loss of sense of time. This often happens during a flow experience, in which time often seem to pass faster than it does. Csikszentmihalyi argues that the distorted sense of time might be a by-product of the other components and not a major component of enjoyment itself, but he adds that freedom from time can boost the flow experience (Csikszentmihalyi, Flow: The psychology of optimal experience, 1990).

The autotelic experience is described as an experience so intrinsically rewarding that the activity is an end in itself. An activity is considered autotelic if it is done for its own sake rather than for the expectation of a future benefit. He adds that few activities are solely autotelic, but instead a combination of autotelic and exotelic (only done for external reasons). The difference between the two is that "*when the experience is autotelic, the person is paying attention to the activity for its own sake; when it is not, the attention is focused on its consequences*." (Csikszentmihalyi, Flow: The psychology of optimal experience, 1990, p. 67).

A flow experience can, as presented earlier, be achieved in almost any activity. This includes sports, creative fields (e.g. music and dance), science, and games (Csikszentmihalyi, 1975; 1990). In addition to this, Csikszentmihalyi states that "*play is the flow experience par excellence*" (Csikszentmihalyi, 1975, p. 37), though he argues that playing a game does not guarantee a state of flow.

The work of Csikszentmihalyi was based on enjoyment in general activities. The theories have later been expanded upon and applied to games.

In a study of uses and gratification of media enjoyment, Sherry (2006) used video games and flow to explain the engagement that players experience during gameplay but argues that media can also be used to prevent flow experiences, depending on the balance between abilities and challenges. Sherry also reports that video games frequently are likely to create flow states as they contain concrete goals and manageable rules, actions that can be adjusted to the skill level of the player, provide clear feedback in terms of scores or progression, and have abundant visual and auditory information that helps concentrate on the task at hand, i.e. the game (Sherry, 2006).

With the flow experience being such a core aspect of enjoyment and so applicable to video games, it was eventually adapted into a game specific model, the GameFlow model, by Sweetser & Wyeth (2005). Whereas the original flow framework was applicable to any activity, the purpose of the GameFlow model was to measure enjoyment in games specifically. The GameFlow model is defined by 8 core elements: concentration, challenge, skills, control, clear goals, feedback, immersion, and social. Most of the core elements are directly related to the components of flow. The social element, however, is unique to the GameFlow model. Sweetser & Wyeth state that social interaction is not a component of flow as it is not directly related to the task at hand and it can even break immersion and the fantasy of living in the game world.

However, social interaction is a large part of enjoyment of games and can even create enjoyment in games that individuals do not like. Social interaction can be supported through competition or cooperation and is enhanced through in-game chats or online boards.

The GameFlow model provides design guidelines for achieving enjoyment in a video game. In total, the model consists of 36 elements related to the eight core elements. The model was evaluated using expert reviews of two real-time strategy games (Sweetser & Wyeth, 2005).

Flow was thus defined as the optimal experience. In its base form, a flow experience was described as an experience where the challenge of an activity matched the skill levels of the player, much like the definition of challenge-based immersion by Ermi & Mäyrä (2005). Through analysing flow, however, it became clear that flow was a much more complicated psychological construct that was made up of nine components.

Flow is an interesting theory for defining player experience as it is a definition of the optimal experience. In addition to being an established theory for evaluating an experience, the flow theory could be used to guide the design of HeartBee, i.e. through designing for the nine components of flow.

This concludes the definition of player experience chapter. In the following chapter I investigate how player experience can be assessed through various measurements. The purpose is to find a theory that can assess the player experience of HeartBee.

2.3 Measuring player experience

Whereas the previous section focused on defining player experience, the purpose of this section is to provide a review of what methods are used to measure and quantify player experience. The focus of this section is on measuring the constructs identified in the previous section, immersion, continuation desire, and flow, but will also include general purpose evaluation tools.

In general, player experience measurements are divided into two categories - subjective measurements and objective measurements.

2.4 Subjective measurements

In this section I review general player experience measurements. The focus is on two measurements - the Game Experience Questionnaire (GexpQ), and the Game Engagement Questionnaire (GengQ).

The GexpQ was developed by IJsselsteijn et al. (2007) as a multi-measure approach of a game experience. The GexpQ tries to capture player experience as a seven-dimensional experience. The seven dimensions are: sensory and imaginative immersion, tension, competence, flow, negative affect, positive affect, and challenge (IJsselsteijn, et al., 2008). The GexpQ contains three modules, all of which are meant to be administered right after playing the game.

The authors of the GexpQ state that five items are needed per dimension to get a robust assessment (IJsselsteijn, Kort, & Poels, 2013). As such, the core module is a Likert scale containing 33 items. Each item is a statement, e.g. "*I felt happy*", that is rated on a five-point scale from not at all (scored 0) to extremely (scored 4). The core module is followed by the social presence module (17 items), which is followed by the post-game module (17 items).

The authors have also adapted the GexpQ to an in-game version which is a concise version of the core module. The in-game version contains 14 items. The in-game version is to be administered at multiple points during gameplay to assess the gameplay experience at different points in time (IJsselsteijn, Kort, & Poels, 2013).

In a review of the GexpQ, Norman (2013) argues that while the questionnaire seems reasonable and applicable in studying player experiences with video games, it will have to be altered for different experiences - especially experiences that do not involve a narrative as well as for non-competitive experiences. Furthermore, he criticises the authors for a lack of empirical evidence on the reliability and validity of the GexpQ (Norman, 2013).

The GengQ is a questionnaire developed by Brockmyer et al. (2009) with the purpose of creating a general, reliable measurement of game engagement in violent video games. The GengQ is a questionnaire that aims to include most descriptions of the multifaceted word engagement. The GengQ is thus based on the constructs absorption, flow, presence, and immersion. The GengQ contains 19 Likert items formulated as statements, e.g. "*I feel like I just can't stop playing*", with possible answers ranging from no to yes.

Like the GexpQ, the GengQ is supposed to be administered immediately after playing the game.

The GengQ questionnaire was iteratively designed and the reliability of the questionnaire was validated internally, making it empirically "stronger" than the GexpQ questionnaire.

The purpose of the GengQ, however, was to develop a "*measure of engagement in playing video games that should be useful in assessing the potential impact of playing video games, particularly violent games*" (Brockmyer, et al., 2009, p. 624). In his review, Norman (2013) argues that the purpose of the scale is not necessarily to measure the player experience, but rather to measure the impact of violent games (in this case, on children) (Norman, 2013). In a response to the review, the authors note that the scale would have to be expanded to capture more common engagement experiences that are not related to violent video games (Fox & Brockmyer, 2013).

The GexpQ and the GengQ are therefore both promising player experience measurements as they include a multitude of psychological constructs that define the player experience. Their shortcomings, however, is their focus on narrative elements (GexpQ) and violent video games (GengQ). Neither of those aspects are present in HeartBee.

2.4.1 Measuring continuation desire

Schoenau-Fog (2011a) argues that a fundamental requirement of an interactive experience is the desire to continue the experience. For evaluating such experience, he proposes the Engagement Sample Questionnaire (ESQ) as well as an intrusive approach, both of which are based on the Player Engagement Process as well as the Experience Sampling Method (Schoenau-Fog, 2011b). In general, the ESQ aims to capture a number expressing the players' desire to continue as well as a qualitative evaluation of why the player wants to / does not want to continue. The qualitative evaluation aims to capture the four main components of the Player Engagement Process (objectives, activities, accomplishment, and affect). Schoenau-Fog (2011a) divides the ESQ into four parts:

The first part is concerned with the demographics of the player and is presented before playing the game.

The second part is concerned with the pre-experience motivation for starting the application. During this part, the player is asked to rate the desire to start the experience as well as the motivation for starting it (the objectives).

The third part is an intrusive method that is repeated several times during the experience in which the user is interrupted in order to evaluate the desire to continue the experience, what makes them want to / not want to continue (objectives) as well as what they were doing in order to achieve their objective (activities), and what they were feeling during the experience (affect).

The final part is concerned with the post-experience desire to try the application again as well as the reasons for doing so / not doing so (objectives).

The complete ESQ (not including demographics) contains 18 questions, 2 pre-experience, 8 during the experience, and 9 post-experience.

In follow-up work, Schoenau-Fog et al. (2012) modified the ESQ towards a simpler and shorter version that could be used as part of an iterative design process. This version is referred to as the basic game ESQ. The basic game ESQ was only concerned with the level of continuation desire as well as the reasoning for that level. In comparison, the basic game ESQ contains 2 questions pre-experience, 2 questions during experience, and 2 questions post-experience.

For both the ESQ and the basic game ESQ, the desire to continue playing is evaluated as a 7-point Likert item, ranging from strongly disagree (scored -3) to strongly agree (scored 3). The evaluation is based on a statement, e.g. "*Please indicate the extent to which you agree or disagree with this sentence: "I want to start the game".*".

The reasoning for wanting to continue is presented as a question with open-ended answers, e.g. "*What makes you want/not want to continue*?".

The basic game ESQ was evaluated during the development of the crowd game "Space Bugz!". The game was tested during five iterations, but respondents were not interrupted during gameplay and only answered the pre-experience and post-experience parts of the basic game ESQ. For each iteration they could then compare the results with previous iterations to identify how the changes made during the iteration affected player engagement (as measured by continuation desire). This allowed the developers to roll the application back to a previous version if upgrades to the game turned out to be downgrades. In general, it is concluded that the basic game ESQ can be used both as a quantitative measure of how good a design is at a certain point in time as well as a qualitative measure for issues in the design as well as engagement and disengagement triggers (Schoenau-Fog, Birke, & Reng, 2012).

In addition to being used as a player experience tool in games, a variation of the ESQ, namely the focus on desire to continue playing, has been used to quantify students' motivation to keep studying (Schoenau-Fog, Reng, & Kofoed, 2015).

Continuation desire can thus be measured using either the ESQ or the basic game ESQ. The short length of the basic game ESQ makes it especially appealing for this thesis due to how easy it is to add to other measurements.

2.4.2 Measuring flow

During his research into a flow experience, Csikszentmihalyi mainly utilized the experience sampling method (ESM) (Csikszentmihalyi, 2014). The experience sampling method is a method that aims to assess what people do, feel, and think at specific moments during their daily lives. The ESM was typically carried out over 1 week. While the method carries many advantages, namely that it identifies experiences as they occur naturally, the task is also rigorous and not suited for experimentally measuring flow.

With that in mind, Jackson & Marsh (1996) set out to develop the Flow State Scale (FSS) which determines a level of flow in an experience from a 36-item Likert scale. The FSS is a scale based on the nine components of a flow experience, each represented by four related questions. The questions are formulated as statements, e.g. "*I felt in total control of what I was doing.*", which are all evaluated on a five-point Likert scale, ranging from strongly disagree (scored 1) to strongly agree (scored 5).

While the FSS provides a rich picture of the flow experience through multidimensional assessment of the nine components of flow, the length of the scale can be off-setting in experiments, especially through multiple measurements of the construct.

To counter this problem, Jackson et al. (2008) developed a short version of the FSS, namely the Short Flow State Scale (SFSS). The SFSS is comprised of nine Likert scale items, one for each component of a flow experience. The formulation of the questions, and the range of the answers, are like those of the FSS. The SFSS was validated internally and proved to provide an acceptable model fit and reliability (Jackson, Martin, & Eklund, 2008).

The SFSS, as well as variations of it, has seen many usages in the literature for evaluating flow in computer games (Tian, et al., 2017; Engeser & Rheinberg, 2008; Brom, et al., 2014) , as well as for other activities such as piano playing (Manzano, Theorell, Harmat, & Ullén, 2010).

Flow can thus be measured through an extensive questionnaire, FSS, or a brief questionnaire, SFSS. Assessing flow through the SFSS is a promising direction for this thesis as it assesses a very centric part of player experience in relatively few questions. At the same time, the SFSS was not designed for games specifically which means it easily scales to other tasks, such as a non-game biofeedback application.

2.4.3 Objective measurements

Unlike subjective measurements, objective measurements are not dependent on the introspection of the player but are instead related to objective measurements. One of such objective measurements is biometrics.

Biometrics are generally divided into two categories - physiological measurements and psychophysiological measurements. Physiological measurements refer to cardiovascular measures, electrodermal activity (EDA), electromyography (EMG), respiration, facial expressions, and eye tracking. Psychophysiological measurements generally refer to Electroencephalography (EEG).

Early biometric measurements were focused on placing biometric measurements on the circumplex model of affect, namely the arousal scale for HR and EDA (Russell, 1980). Advancements in the field of biometrics, however, as well as advancement in technology, has led to biometrics being able to measure more complex constructs than arousal, such as immersion and flow.

Jennet et al. (2008) investigated how eye movement was related to immersion by comparing a non-immersion condition with an immersion condition. They found that the participants in the non-immersion condition significantly increased their eye movements over time, whereas participants in the immersion condition significantly decreased their eye movements over time. They correlated self-reported immersion with eye movements and found that the two were somewhat associated (Jennett, et al., 2008).

In a study of psychophysiology and flow during piano playing, Manzano et al. (2010) found that flow was significantly associated with both EMG, HR, and respiration. They found that higher levels of flow were associated with increased respiratory depth, reduced RSA, reduced cardiac output, and increased EMG activity (Manzano, Theorell, Harmat, & Ullén, 2010).

Tian et al. (2017) performed a physiological signal analysis for evaluating flow during a computer game of varying difficulty. They found that flow was associated with increased respiratory rate, increased respiratory depth, moderate HR, moderate HRV, and moderate EDA (Tian, et al., 2017).

While more complex as a measurement, several studies have found correlations between EEG and player experience, e.g. engagement (Leiker, et al., 2016) and flow (Ulrich, Keller, Hoenig, Waller, & Gröna, 2014). The findings were based on correlating the EEG measurements with self-reports of the player experience.

While biometrics seem promising for measuring the user experience, one of their shortcomings is their lack of cause-effect relationship.

Another objective measurement is the time spent playing the game. Nacke & Drachen (2011) define the time spent, and how it is spent, as the temporal dimension. They go on to state that time spent is closely linked to psychological concepts such as flow, immersion, and presence (Nacke & Drachen, 2011).

In another study, Johnson et al. (2016) set out to find the predictors of time spent playing video games. Their findings suggested that playing with others (socializing) was associated with more time spent playing, and that all three elements of self-determination theory (competence, autonomy, and relatedness) were associated with more time spent playing. Their findings, however, were based on a survey in which people had to recall how much time they spent (Johnson, Gardner, & Sweetser, 2016). Wu et al. (2017) argue that time spent is a strong measure of engagement when watching videos (Wu, Rizoiu, & Xie, 2017).

Player experience can thus be assessed through a variety of objective measures. Physiological measurements were introduced as a potential for assessing several psychological constructs. Time spent was introduced as an objective method for assessing engagement.

This concludes the measurement of player experience chapter. In the following chapter I delimit the problem by analysing the theories and how they relate to HeartBee. The delimitation concludes in the final problem statement that I design for and investigate throughout the rest of this thesis.

2.5 Delimitation

In this section, I delimit the research from the initial problem - How can a mobile HRV biofeedback game increase the user experience during biofeedback training? - into a more precise and narrow final problem.

2.5.1 Analysis of player experience measurements

The review of engaging user experience in games let to the term player experience. An engaging player experience was presented as a multifaceted term. The focus of this thesis was on three different constructs: immersion, flow, and continuation desire.

Immersion was described as a metaphor for being submerged in water - the feeling of being totally absorbed and present in the medium. The descriptions for immersion in games were somewhat varying. It was described as an involvement in the narrative and the characters of the game, whereas engagement was described as the involvement with the non-diegetic elements of the game, e.g. scores. It was also described as a three-stage process - from engagement to engrossment to total immersion. Another theory described three types of immersion - sensory, challenge-based, and imaginative.

What was common for all descriptions of immersion was the need for a strong narrative and empathy with the characters in the world. In the case of HeartBee, an endless runner with a lack of story and character development, immersion seems like a bad fit for measuring player experience.

The Game Experience Questionnaire (GexpQ) is an interesting and extensive player experience measurement as it attempts to capture seven dimensions of the player experience, namely the sensory and imaginative immersion, tension, competence, flow, negative affect, positive affect, and challenge. In addition to that, the GexpQ offers an intrusive in-game player experience measurement which can be used as part of an iterative design process.

The GexpQ, however, also has a focus on narrative experiences and lacks empirical evidence for its reliability and validity. On top of that, the questionnaire contains 67 items. This would increase the time needed for the experiment immensely and I therefore deem the questionnaire not fitting for this project.

The Game Engagement Questionnaire (GengQ), on the other hand, contains 19 items that are based on the four constructs absorption, flow, presence, and immersion. Furthermore, the questionnaire has gone through thorough validation which makes it an attractive option. The focus on the questionnaire, however, was on violent games, and it would have to be tweaked to fit other game genres. With HeartBee being the contrary, a game that promotes relaxation, it therefore does not seem to fit as an evaluation tool either.

What Immersion, the GexpQ, and the GengQ have in common is their focus on measuring player experience in games. In this project I am interested in this measurement, but I am also interested in measuring player experience in a non-game application.

Using biometrics for measuring player experience is another option due to their objectiveness and recent advances suggesting that they can capture more complex psychological constructs such as flow. One can argue, however, that the active changes in physiology that occurs during HRV biofeedback training would provide unreliable measurements with biometrics. The measurements would then be a result of the active participation of the participant (e.g. raised HRV due to deep breathing) rather than a result of psychological constructs. Biometrics are therefore also deemed as not fitting for measuring player experience for this project.

A flow experience was described as an optimal experience - the experience of being in the zone. It was constructed by nine components - A challenging activity that requires skills, the merging of action and awareness, clear goals, clear feedback, concentration on the task at hand, the paradox of control, the loss of self-consciousness, the transformation of time, and the autotelic experience. The flow experience could be captured through the flow state scale (FSS), a 36-item scale with 4 items for each component. A shorter version of the FSS, the Short FSS (SFSS) with only nine items, one for each component of flow, had been used to capture the same flow experiences with acceptable reliability. Furthermore, the SFSS had commonly been used to capture flow experiences during computer play as well as other activities.

Due to its reliability and its common usage in computer games, as well as other activities, flow as a measurement for an engaging player experience fulfils the requirements put forth in this thesis. Furthermore, the short version is especially useful as it measures the experience in only nine items.

Continuation desire was described as another measure of engagement - a measure of a player's desire to continue playing the game. The desire to continue playing was described to be related to emotions like satisfaction, enjoyment, and fun, as well as experiences like immersion, presence, and flow.

The desire to continue playing the game can be captured through the Engagement Sample Questionnaire (ESQ) as well as the basic game ESQ. The power of using the ESQ, and especially the basic game ESQ, is that it captures the desire to continue playing the experience in very few items. In addition to this, they both allow for obtaining qualitative data, e.g. why a player wants to/does not want to continue playing. This contrasts with all the previously mentioned methods that only capture quantitative data. This makes the ESQ and the basic game ESQ very prominent during an iterative design process. Furthermore, the ESQ has been used to capture a desire to continue outside of gaming scenarios, which makes it applicable for a non-game HRV application.

As such, continuation desire as a measurement of an engaging player experience is another measurement that fulfils the requirements of this thesis. The basic game ESQ is especially valuable for the iterative design of the game.

Time spent playing was mentioned as an objective measurement that was closely linked to engagement, flow, immersion, presence, as well as self-determination theory. The literature behind time spent, however, was limited and focused on players recalling their time spent playing. Due to its objectivity and completely non-intrusive nature, time spent is another parameter that could easily be included in the measurement of player experience.

In summary, measuring player experience, in this project, can be boiled down to measuring flow, continuation desire, and time spent. The three methods can easily be combined and measured during the same experience due to their shortness in length (flow and continuation desire), as well as the passive unobtrusiveness of measuring time spent. In addition, using several player experience measurements can give a more precise and well-rounded evaluation of the player experience compared to just using one measurement. This is also seen in how the GexpQ and the GengQ are constructed.

Finally, to my knowledge, flow, continuation desire, and time spent have not been correlated and reported in the same experiment. Measuring the three constructs and correlating them could give insight into their relationship with each other and be valuable for future definitions of player experience.

The second part of the delimitation deals with finding a non-game HRV biofeedback application that HeartBee can be evaluated against.

2.5.2 Analysis of available HRV biofeedback applications

This section aims to provide a review of the currently available HRV biofeedback applications on the market. The focus is on identifying an app that can be used to evaluate HeartBee against. During the review, I noted that not a single HRV biofeedback app was developed as a game, which shows a big potential for the application developed during this thesis.

The initial review was done by searching for “Heart rate variability” on AppCrawlr (TipSense LLC, 2018) on August 10th, 2018. This returned a result of 31 applications. The first step was to sort out the applications that did not measure HR, but just mentioned it as part of their description. This yielded a sample of 24 applications. A further reduction of the sample was done by removing applications that focused on HR during physical exercise rather than HRV and resting. This reduced the sample to 16 applications.

Finally, HRV can be measured through either external devices (e.g. ear clip sensors or chest straps) or by placing the finger on the camera of the smartphone. Due to the lack of monetary

Table 2 The scores of four apps based on the four requirements.

	IF	BG	UT	Score
SCA	No	No	No	0
HS	No	No	No	0
HRCP	Yes	Yes	(Yes)	2.5
SRM	Yes	Yes	No	2

IF = Immediate Feedback, BG = Breathing Guidance, UT = Unlimited Time resources allocated to this thesis, applications that required external devices for recording HRV were excluded as well. This yielded a final sample of four applications - Stress Check by Azumio (SCA), Heartservice (HS), HeartRate+ Coherence Pro (HRCP), and Stress Releaser Meditation (SRM).

The four applications were then analysed based on three criteria: It should provide immediate HRV feedback (e.g. SDNN, coherence), it should provide a guidance for breathing, and it should allow the user to train for an unlimited amount of time. The applications are compared in Table 2 and are associated with a final score for how many of the criteria they fulfil.

From the table it is evident that the only application that fulfils all requirements for the experiment is HRCP. It scored a 2.5 out of 3. The free version limits sessions to 90 seconds whereas the full version does not limit the time. HRCP is also considered a relatively popular application with more than 10.000 downloads and a user rating of 3.5 out of 5.

HRCP was therefore purchased and is chosen as the non-game HRV biofeedback application that HeartBee is compared to. An alternative to buying HRCP would be to develop a non-game HRV biofeedback application. Due to the limited time-frame of this thesis, as well as to the noise that could be introduced by comparing to a non-verified application, doing so was dismissed. The user interface (UI) of HRCP is shown in Figure 4.

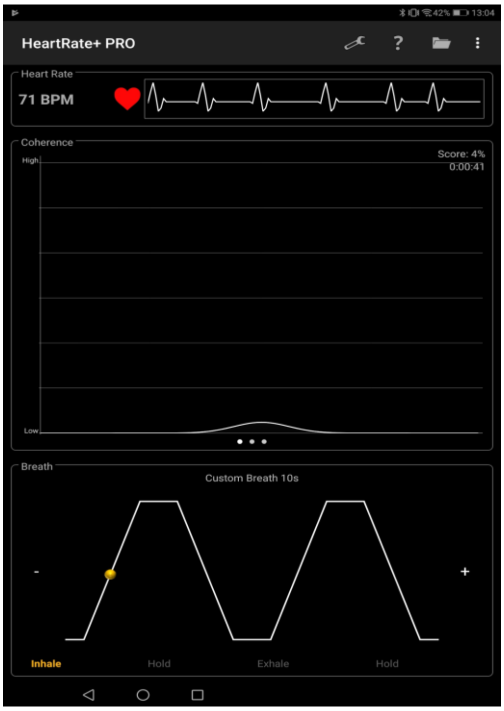


Figure 4 The UI of HRCP.

2.5.3 Final problem statement

The purpose of this thesis is to evaluate the player experience of HeartBee and compare it to HRCP. The previous chapter delimited the measurement of player experience to flow, continuation desire, and time spent. The three measurements were chosen as they could easily be combined and applied to HeartBee to give a broader picture of the player experience.

With all the above in mind, the initial problem has been refined into the final problem statement:

> *To what extent can a HRV biofeedback game elicit higher levels of flow, higher levels of continuation desire, and more time spent training compared to a non-game HRV biofeedback application?*

The final problem statement leads to three main hypotheses:

- **H1**: The HRV biofeedback game elicits higher levels of flow than the non-game HRV biofeedback application.
- **H2**: The HRV biofeedback game elicits higher levels of continuation desire than the non-game HRV biofeedback application.
- **H3**: The HRV biofeedback game leads to more time training than the non-game HRV biofeedback application.

In addition to the main hypotheses, it was stated that flow, continuation desire, and time spent could be related to each other. This leads to three secondary hypotheses:

- **H4**: Higher levels of flow are related to higher levels of continuation desire.
- **H5**: Higher levels of flow are related to longer time spent training.
- **H6**: Higher levels of continuation desire are related to longer time spent training.

At this point, I have delimited the research into a final research question which describes the comparison between an HRV biofeedback game (HeartBee), and a non-game HRV biofeedback application (HRCP) from a player experience perspective.

This concludes the background and motivations behind this thesis. The theory of both HRV and player experience have been described through research and have been delimited to a final problem statement with a set of hypotheses. The following chapters of this thesis focus on the chosen methodology of the thesis, how the design was carried out - both from an interaction design perspective and with the purpose of fulfilling the requirements of this thesis, as well as how the design was implemented.

3 METHODOLOGY

In this section I describe and argue for the methodology chosen to evaluate the problem statement and hypotheses presented in the delimitation section of the background chapter.

The problem statement of this thesis is associated with a set of hypotheses that I want to confirm through the experiment. The hypotheses are all based on established theories within player experience research and as such, the goal of this thesis is to investigate these hypotheses. The overall methodology is therefore confirmatory research. In the following section I describe the experimental design.

I established that the basic game ESQ, flow, and time spent could be combined as a holistic approach for measuring the quality of a player experience. Furthermore, it was established that the basic game ESQ was a valuable tool for an iterative design process. I therefore argue that the methodology used throughout this thesis is mixed methods (Bjørner, 2015). Mixed methods research is the type of research where the researcher gathers both qualitative and quantitative data. As described in the delimitation section, a quantitative measure of a player experience provides a number for the experience, but it does not explain why the experience was rated such. Mixed methods are therefore able to provide breadth and depth of the understanding of the findings.

The overall methodology of the experiment has now been established. In the following section I describe and argue for the experimental design chosen for this thesis.

3.1 Experimental design

In a systematic review of the measurement of engagement in games, Boyle et al. (2012) found that most study designs (9 out of 13) utilized a quasi-experimental design when it came to measure the subjective experience of engagement. In comparison, only two used a randomized controlled trial (RCT) design. The authors state that an RCT design should be the experiment design of choice (Boyle, Connolly, Hainey, & Boyle, 2012).

Two design methods that both fulfil the RCT design were considered for evaluating this thesis: A between groups (independent measures) design and a within groups (repeated measures) design.

The independent measures design is in its simplest form a design in which participants are allocated into two groups - a treatment group and a control group. Scores can then be compared between the two groups (Field & Hole, 2003). Allocation of the participants should be random to fulfil the RCT requirement. This is also known as a parallel-group design.

The experiment of this thesis lends itself nicely to such a design as the non-game HRV biofeedback system could be considered a control condition whereas the game could be considered the treatment condition. Field & Hole argue that there are two main disadvantages to an independent measures design: One is that it is expensive in terms of time, participants, and effort. The second is that it is less likely to detect effects of the experimental manipulations, e.g. due to noise coming from random differences in the two groups.

The repeated measures design is an experimental design in which each participant is exposed to all conditions of the experiment. The scores are then compared within the group instead of between the groups. The strength of this method is that one condition can be a baseline measurement. In comparison to the independent measures design, the repeated measures design is cheaper and is more likely to detect effects of experimental manipulations, thanks to the baseline measurement. In fact, Field & Hole state that "*where it's feasible, use a repeated-measures design*" (p. 97). They identify two conditions that can make a repeated measures design infeasible: One being the risk of carry-over effects from one condition to the other and the second one being that the conditions need to be reversible, i.e. participating in one condition does not have irreversible effects that makes the participant unable to take part in another condition.

Field & Hole define carry-over effects as variations in answers because of systematic effects. These include fatigue, boredom, and becoming better over time. To control for these systematic variations, they argue for a counterbalance approach. In this approach, half of the participants are presented with condition A then B, whereas the other half is presented with condition B then A. By doing so, the researcher is also able to consider the order of presentation as a variable that can be analysed.

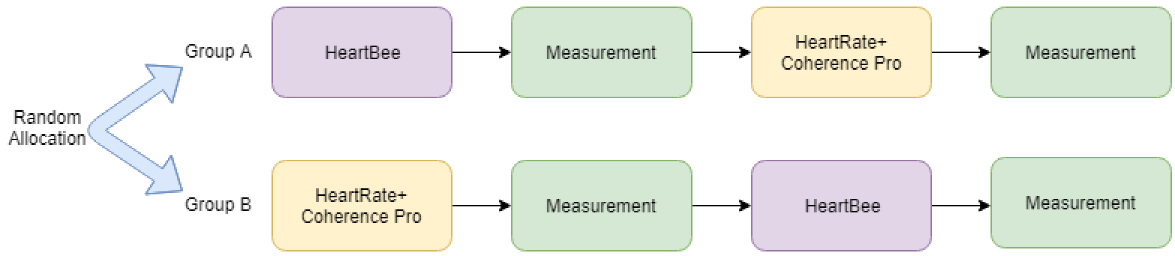


Figure 5 Visualization of the experimental design.

To fulfil the RCT requirement, the participants should be randomly allocated into group A and B. This is also known as a crossover design (Field & Hole, 2003).

I argue that neither the game or the non-game application has an irreversible effect on the other condition - playing HeartBee does not mean that you cannot use HRCP and using HRCP does not mean that you cannot play HeartBee.

I argue that the two conditions that could make a repeated measures design infeasible are not evident in the experiment of this thesis and the experimental design of choice is therefore a repeated measures design with the counterbalance approach. The approach is visualized in Figure 5.

In the following sections I describe how the three dependent variables, continuation desire, flow, and time spent are measured.

3.2 Measuring continuation desire

The measurement of player engagement was described through the theory of continuation desire which was based on the OA3 framework. The ESQ and the basic game ESQ were described as two tools for measuring continuation desire. The extensiveness of the ESQ, regarding the amount of questions posed, makes it less suitable for the combined methods proposed in this thesis. As such, the basic game ESQ is chosen as the measure of engagement. The basic game ESQ contains three statements that the player must consider: "*I want to start the game*", "*I want to continue playing*", and "*I want to try again*". The three questions are administered pre, during, and post gameplay respectively. The three questions are answered on a scale from 1 (strongly disagree) to 7 (strongly agree). Each question is followed by the qualitative part of the basic game ESQ, e.g. "*what makes you want/not want to try again?*".

For the design phase of this thesis, a full version of the basic game ESQ is chosen as to provide a perspective into the player experience of the game during development. Additionally, the design phase is not focused on measuring either flow or time spent, making it more suitable to administer the full version of the basic game ESQ.

For the main evaluation of the thesis, however, I deem it impossible to administer the full version of the basic game ESQ. Intrusively stopping the experience midway is believed to break a flow experience as players become aware of their surroundings and the experiment. Furthermore, it biases the time spent on the application greatly.

Measuring the motivation for starting the experience is also deemed unnecessary as the players are not approaching the game/application by free will but rather to participate in a study. The post-game evaluation, however, is valuable in getting information about engagement triggers and disengagement triggers, and provides a qualitative analysis of the game, thus keeping the mixed-methods approach. The open-ended responses will be analysed through traditional coding (Bjørner, 2015) and compared to the categories of the OA3 framework.

3.3 Measuring flow

The delimitations chapter suggested that flow could be measured through two scales - the FSS and the SFSS. The strength of the FSS is that it has multiple items for each component of a flow experience whereas the SFSS has only one per component. The strength of the SFSS, then, is that it measures flow through only nine items, compared to 36 of the FSS. The limited length of the questionnaire is essential in a study where multiple questionnaires are utilized. Furthermore, the repeated measures design doubles the amount of questions that each participant must answer, so a reduction where possible is sought after. In addition to that, the SFSS has been utilized by several studies investigating flow in a game. The SFSS has therefore been chosen as the measurement of choice for measuring flow in this thesis. The questionnaire consists of the following nine statements (Jackson, Eklund, & Martin, 2010):

- I felt I was competent enough to meet the demands of the situation
- I did things spontaneously and automatically without having to think
- I had a strong sense of what I wanted to do
- I had a good idea about how well I was doing while I was involved in the task/activity
- I was completely focused on the task at hand
- I had a feeling of total control over what I was doing
- I was not worried about what others may have been thinking of me
- The way time passed seemed to be different from normal
- I found the experience extremely rewarding

Respondents are asked to rate each statement from 1 (strongly disagree) to 5 (strongly agree).

As explained in the previous section, measuring flow is not of interest during the design phase of the thesis. It is therefore only measured during the final experiment.

3.4 Measuring time spent

Time spent is an objective measure and does therefore not warrant a methodology. There are, however, two variables that need explanation: when does the measurement begin, and what is the maximum amount of time allowed.

The time spent during HeartBee should be measured from when the participants start playing the game (tutorial included). Any setup (e.g. setting up the sensor and registering a name) should be excluded, as it is not part of time spent on the actual application.

Time spent during HRCP is measured from when the participant starts the measurements after an introduction to the application.

The time spent should be limited to reduce the time spent on the experiment. The maximum allowed time spent is 15 minutes per application as this allows an experiment to complete in less than an hour. Most HRV biofeedback training applications report that they should be used for at least 10 minutes, though some report 20 minutes of training during their studies, e.g. (Paul & Garg, 2012). Limiting the time to 15 minutes thus captures the minimum requirement while allowing more time to engage/disengage with the applications. It should be noted that participants should not be told about the actual time cap, as that can give them a sense of wanting to play until the experiment ends.

The final design of the questionnaire is divided into three parts: demographics, measuring flow, and measuring continuation desire. Demographics are concerned with the gender, age, nationality, and occupation of the participants. The other parts have previously been described. Part two and three, flow and continuation desire, are repeated twice, once for each condition. Two version of the questionnaire have been designed - one for the HRCP first group and one for the HeartBee first group. Both versions are designed in Google forms and are supposed to be filled out from the device that the participants use for testing. The complete questionnaires can be found in the digital appendix.

This concludes the methodology behind evaluating this thesis. A counterbalanced repeated measures design was chosen for evaluating the main hypotheses. Flow will be measured through SFSS and continuation desire will be measured using the post-game part of the basic game ESQ. Time spent will be measured from when participants actively start using HRCP or when they start the tutorial in the case of HeartBee. Time spent will be limited to a maximum of

15 minutes to keep the experiment within a fair time frame while still allowing time for the participants to identify disengagement triggers.

Continuation desire will also be used for guiding the iterative design process of HeartBee. The iterative design process of HeartBee will be presented in the following chapter.

4 Design

This chapter aims to describe the design iterations and final design of the HRV biofeedback game, HeartBee. The design guidelines will be propelled from two sources - one source being the previous iteration of HeartBee, for which qualitative data (in relation to continuation desire) was collected to help guide the design for future iterations. The other source is the final problem statement and the requirements for testing it. For readability purposes, the previous iteration of HeartBee is referred to as the pre-thesis iteration.

An iterative design process was used in the design process as creating a perfect design in the first try is near impossible (Nielsen, 1993). The idea of the iterative design process is that it is a cyclic process in which designers establish requirements for the design, develop a design that fulfils those requirements, and evaluate it on the users. This can then establish new design requirements and then the cycle repeats. The iterative design process, also known as the basic interaction design cycle, is shown in Figure 6.

The iterative nature of the basic ESQ lends itself well to evaluating products from both a usability and player experience standpoint, as was the case for (Schoenau-Fog, Birke, & Reng, 2012).

In the following subchapter I briefly describe the iterations that led to the initial design of this thesis as well as the results of the pre-thesis final test, as to provide a starting point of the present design cycle. An elaborate description can be found in (Brejl, 2016).

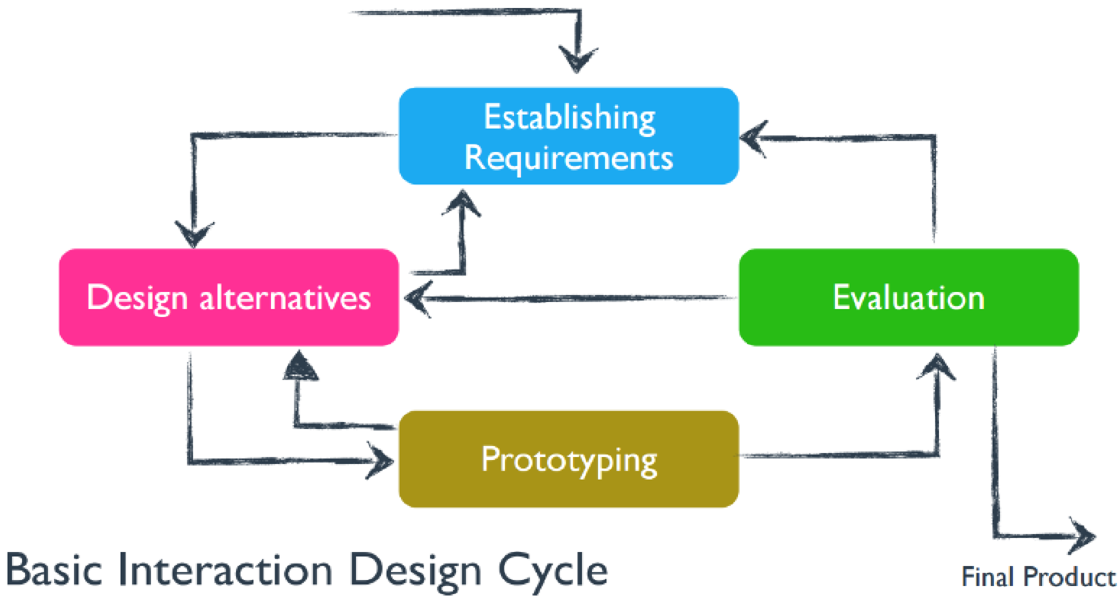


Figure 6 The basic interaction design cycle (Gelineck & Wang).

4.1 Summary of the pre-thesis design cycle

In the pre-thesis design cycle, HeartBee saw a transformation from a simple “sinusoid runner” to a fully-fledged endless runner game. The initial design was guided by design requirements from HRV research, e.g. that the game should provide a real-time visualization of the HR, the game should be based on breathing following a 10 second rhythm, the game should incentivize players to achieve a sine-wave like pattern in their HR, and it should promote positive emotions. From player experience research, the design was guided by e.g. a need for clear goals and a matching of skills and challenge.

A focus group consisting of four Danish university students (one female) ranging in age from 23 to 26 years of age was recruited to help generate ideas for the game - both from a mechanics perspective and from a theme perspective. They suggested that the aesthetics could be based on a field landscape (as in the game Flower) and that the game should be penalty free and show progression.

Through an initial user evaluation, it was found that the lack of penalty could make the game boring, leading to a re-introduction of challenges. Furthermore, it was discovered that participants tended to hyperventilate to perform better in the game. In general, there was a lack of explanation of the controls of the game.

The final design of the pre-thesis game introduced a tutorial level that was played before the real game. The tutorial level explained the controls and how to control the HR through breathing. A screenshot of the pre-thesis tutorial level is shown in Figure 7.

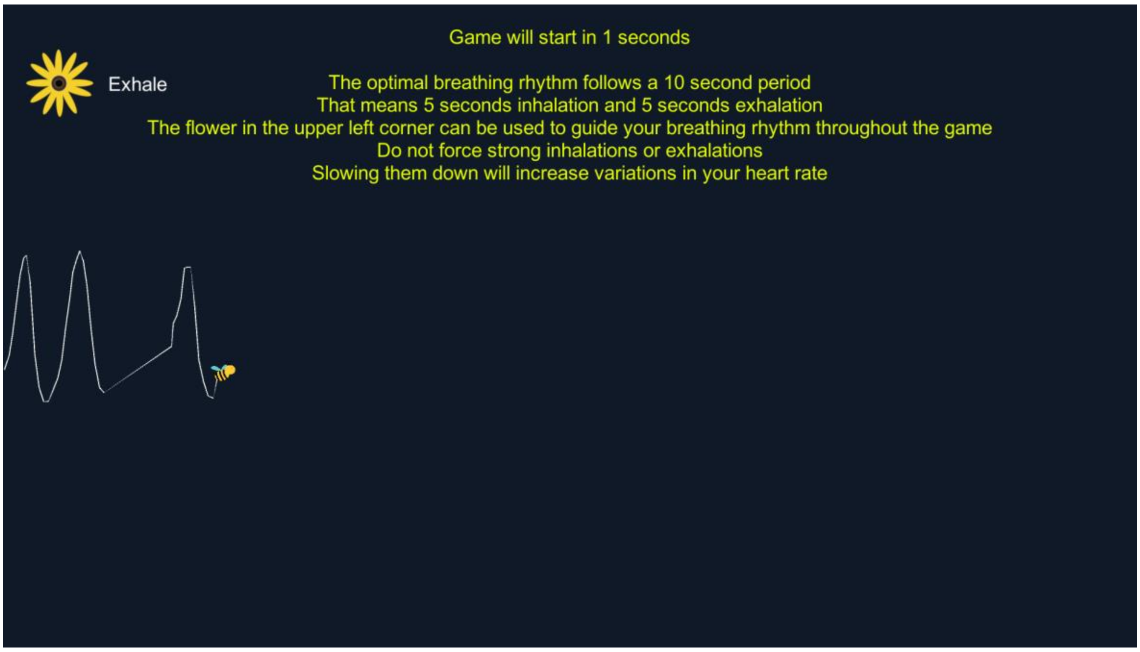


Figure 7 Tutorial level from the final pre-thesis design of HeartBee.

The final evaluation of the pre-thesis game was done as part of the experimental test in which HRV was compared between two conditions. The game was played for 10 minutes for all participants. The findings revealed that disengagement triggers were that the play time was sufficient (i.e. they got bored), the specified breathing rhythm was uncomfortable, the game was too difficult, or the game was too repetitive. The findings, however, were based on only the last question of the basic game ESQ, i.e. "*do you want to try again*" and a binary measure was used, i.e. yes or no.

Engagement triggers of the game revealed to be that they wanted to beat the high score, wanted to become better at controlling their HR, that they had a relaxing experience, and that they enjoyed the controls. Regarding the audio-visual experience of the game, nine out of the 39 (23%) people who played the game stated, unmotivated, that they enjoyed the sounds, animations, and/or graphics, e.g. "*I really liked the graphics and animations in the game. It motivated me to play longer. Also the music was nice which also motivated me*" (Brejl, 2016). One participant stated that he did not enjoy the aesthetics but elaborated that the lack of enjoyment was aimed towards the Mario pipes in the game. A screenshot of the pre-thesis final version of HeartBee is shown in Figure 8.

Finally, observations during the experiment showed a general discomfort for wearing an ear clip sensor. The sensor also provided unreliable results with a lot of noise, which might have affected players' sense of control during the game. Furthermore, the whole setup required an Arduino and several wires, making it less mobile, more intrusive, and unable to run on a mobile device.

Figure 8 Main level from the pre-thesis final design of HeartBee.

4.2 Improving the design

With the pre-thesis iteration in mind, and with the focus of this thesis being on mobile devices, the main design requirement for the first iteration was to develop the game with a non-intrusive measure of HR.

4.2.1 First iteration

For a non-intrusive measurement of HR, three solutions were considered: ECG from chest-straps, camera-based PPG (as used in HRCP), and PPG received from wearables (e.g. smartwatches).

ECG measurements are generally reported to be the most precise measurements of HR and several consumer-friendly products exist, e.g. the Polar H10 heart rate sensor (Polar Electro, 2018). With the reliable measurements of chest-straps, however, comes a level of intrusion as the players will have to wear a chest strap whenever they want to play the game. In addition to that, Aalborg University Copenhagen (AAU CPH) was not in possession of such chest-strap and it was therefore dismissed as a solution for the project.

Using camera-based PPG is a very interesting solution as it is completely non-intrusive, albeit the player is not allowed to move the finger during play. An implementation of such system is presented by Pelegris et al. (2010). They report that it achieves a relatively high accuracy, when compared to a pulse oximeter. They, however, also report that the algorithm is very sensitive to movement. Due to this sensitivity, and due to an implementation of the algorithm being out of the scope for this project, this solution was also dismissed.

This brings me to the final solution that was chosen for this thesis: measuring HR from a wearable. Specifically, two wearables were considered for this thesis - The Samsung Gear S3 (Samsung, 2018) and the Empatica E4 wristband (Empatica, 2018). The Gear S3 was provided by the Samsung Media Innovation Lab (SMILE) at AAU CPH whereas the E4 was provided by the Augmented Cognition Lab at AAU CPH. The initial idea was to utilize the Gear S3 for measuring HR, but an unclear documentation on whether interbeat intervals (IBI) could be extracted from their API, as well as a lack of evidence on the validity and reliability of the measurements led to ultimately dismissing that idea as well.

The E4, in contrast, has been validated against an ECG (McCarthy, Pradhan, Redpath, & Adler, 2016). The Empatica E4 wristband was thus chosen as the device for measuring HR for HeartBee. From this point on, the E4 is referred to as the sensor as well.

In addition to an updated sensor, the first design iteration of this thesis also included an updated version of the control scheme. The old control scheme had the bee moving at a constant speed towards its next position (based on the HR). The new control scheme considered the movement of the bee binary - i.e. either your HR was rising or falling. During a rising HR, the bee would move upwards, while during falling the bee would move downwards. This control scheme was not as much a direct visualization of the HR, as was one of the original design requirements, but was instead an approximation of the visualization. The idea behind this new control scheme was that the movement of the bee would feel smoother and participants would get a larger sense of control. The first iteration of the game was evaluated as a computer application due to the challenges of developing and especially upgrading this application on a mobile device (see 5.2).

The new sensor and the new control scheme was evaluated on three people, all male, ranging in age from 26 to 29 years old, two of which were graduates with a master's degree in Medialogy and a specialization in games.

The evaluation was performed using the basic game ESQ with a few modifications to focus on the experiment at hand:

The first part, the motivation for starting the game, was presented AFTER the participants had played through the tutorial. I argue that this gives a better representation of their motivation to try the game as it would otherwise be based on them being recruited for the experiment.

Secondly, the second and the third part (in-game and post-game) evaluations were repeated two times, once for each control scheme.

In addition to evaluating the new elements of the game, the purpose of the basic game ESQ was to identify engagement and disengagement tiggers in the game. By utilizing all three parts of the basic game ESQ, I hypothesized that the findings would be more precise and in depth than those gathered in the final test of the pre-thesis project.

The participants elicited a high level of motivation for playing the game ($M = 6$). They reported that their motivations were based on being intrigued by an alternate controller, e.g. "*I feel like I understand the controls now and I am very intrigued to try it out! Its exciting to play a game with an alternate controller."* (respondent 1, M, 26). Another respondent reported that it was the unknown that motivated him to start playing. It was also reported, however, that the tutorial "*looked boring and confusing*" (respondent 2, M, 29), which made one participant rate the motivation as 5.

The results of the three participants' desire to continue playing and to try again are reported in Table 3. Possible values ranged from 1 (strongly disagree) to 7 (strongly agree).

The desire to continue playing and the desire to try again was generally rated high amongst the participants, though slightly higher for the second control scheme. Coding of the answers, however, revealed that this increase in continuation desire was not because the control scheme was better, but because the sensor, in the first playthrough, was very unresponsive, e.g. "*I really want to try the game, but it does not work*" (respondent 2, M, 29). One participant, however, reported that the sensor felt more responsive during the second playthrough and that he felt like he could do better because of that while another one reported that the second iteration felt easier to control. Both reports, however, can be due to more consistent readings from the sensor.

The players generally reported that their reason to want to continue playing were that they felt like they were getting control over the bee. Furthermore, they reported that they wanted to beat the high score and beat their friends.

Finally, the players reported that they wanted to play again to beat the high score and because they liked the audio-visual landscape of the game. Additionally, one player identified that the

Table 3 Reported continuation desire as a function of control scheme.

	PTCS	NCS
I want to continue	5	6
I want to try again	5	6

PTCS = Pre-thesis control scheme

NCS = New control scheme

UI, both during tutorial and during game, was confusing and could be upgraded. Specifically, he replied that "*The UI which informs of how to breathe should be more clear and more centered in the experience*" (respondent 2, M, 29). The same player also identified the Mario-pipes as being inconsistent with the aesthetic style of the game.

The participants were debriefed about the experiment after completing the basic game ESQ. During the debriefing they were encouraged to provide further feedback. One respondent added that there was too much text in the tutorial and that he because of that did not read the whole thing.

All participants were asked specifically to elaborate on their thoughts of the new control scheme. One did not notice the difference, whereas the two others indicated that the game could easily become too easy with the "autopilot" and that they did not feel as strong a sense of control over the bee. It was, however, a nice addition when the sensor lost connection as it gave the sensor time to respond.

The design requirements derived from the first iteration were then that the high score system was a central part of the experience and should be expanded upon, and the UI should be upgraded with the breathing guidance being more centered. Another requirement that arose from the test was that a mixture of the two control schemes would be beneficial such that the player had a sense of feedback even after the sensor lost connection.

4.3 Second iteration

Whereas the first iteration was a classic example of the design cycle, the second iteration can be considered a multitude of design cycles evaluating isolated parts of the game. An extensive implementation period, as well as the limited time scope of this thesis, made it impossible to carry out further tests in the sense of using the basic game ESQ. Instead, isolated parts of the game were evaluated by the two graduates from the first iteration through expert reviews. The following list is an extensive list of what was designed and reviewed during this iteration of the design:

Update the font to be more exciting - One of the reviewers originally expressed that the UI was boring. He argued that this was largely due to the standard font used. The font was therefore upgraded by using the TextMesh Pro (Unity Technologies, 2018) text system instead of the standard text system of Unity. Specifically, most of the text was changed to using the Roboto font by Google which is designed for Android. The score text, however, was changed to using the Bangers font to create more variety. The changes were received positively by the reviewers.

Update tutorial level - The reviewers reported that the tutorial was boring and that there was too much text. Furthermore, there was a lack of focus on the breathing aspect of the tutorial. As previously described, the font was upgraded to make it look more inviting. Additionally, the walls of text were instead turned into short sentences that were presented one at a time.

The tutorial also put a large focus on the flower that guides the breath through centring and enlarging it during most of the tutorial. During the pre-thesis iterations, the flower would expand during exhalation and shrink during inhalation. This was based on the design of an analysed game. This, however, contrasts with how the lungs function where an inhalation expands the lungs. Other visualizations of breathing, like the one found at Google (Google, 2018), are in line with this. The flower was thus changed so that it expanded during inhalation and shrank during exhalation.

The flower is also animated to the upper left corner in the end of the tutorial with accompanying text to remind the player that the flower can always be used to guide their breath. The updated tutorial design can be seen in Figure 9 which illustrates an increased focus on the flower and an example of one of the short descriptive sentences that replaced the walls of text.

Update main level - The UI of the main level was updated to be in line with the previously described UIs. The main level starts, much like the tutorial level, with a large centered flower to guide the player's breathing. It is animated towards the upper left corner after one cycle of breathing. To give the player a chance to get into the game, the obstacles are also moved back accordingly so that the first obstacle occurs after one cycle of breath.

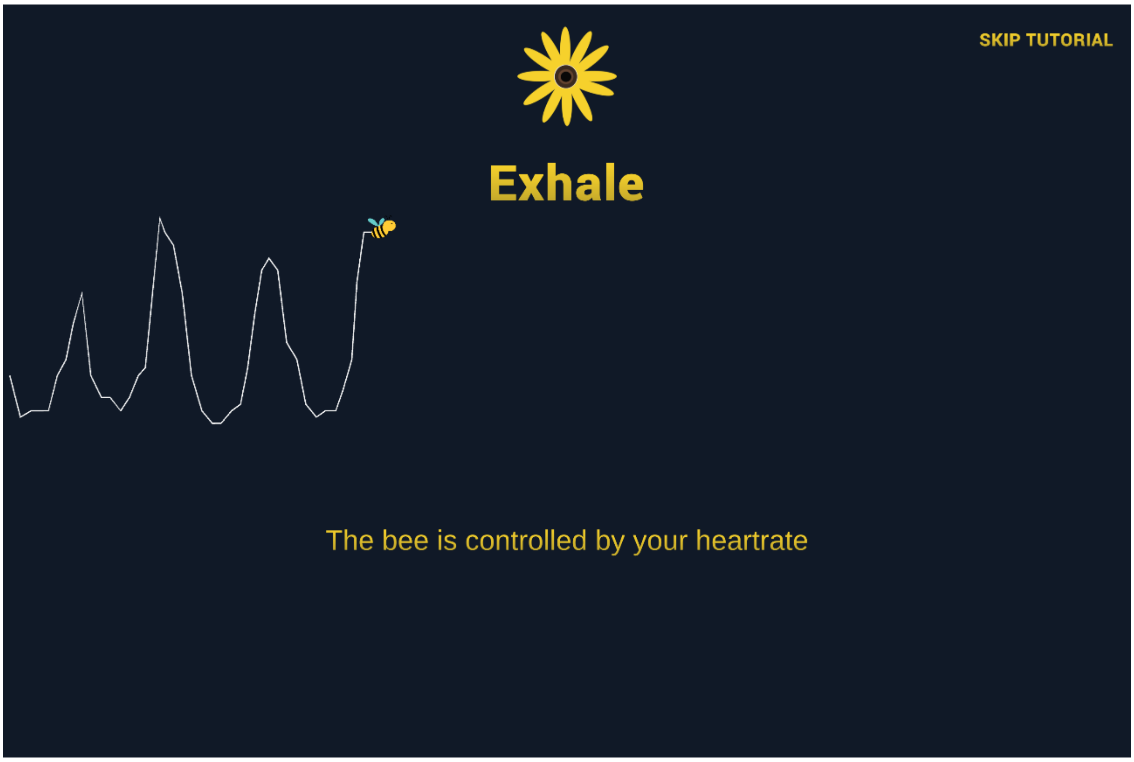


Figure 9 The tutorial level in HeartBee.

The score text of the pre-thesis game provided redundant information such as text stating "*Score:* ", "*Personal best is:* ", and "*Highscore:* ". These redundant texts were removed and only the current score, as a number, was left in the top of the screen. This change is in line with other similar games, e.g. Flappy Bird.

The Mario pipes, also inspired by Flappy Bird, were reported to break the coherence of the aesthetics. The Mario pipes were thus replaced with tree trunks. The updated main level is shown in Figure 10 which also shows the updated score feedback (the number 6 in the upper centre).

Designing the high score system - One of the core aspects for creating engagement in HeartBee is the ability to beat your own score as well as the scores of your friends. The pre-thesis version of HeartBee only visualized the player's current score as well as the best score set on the device. The new high score system is designed to include the player's current score, the player's best score, as well as a leader board with the five best scores. These scores are global and are thus not device specific as in the earlier iteration.

In addition to this, the expert reviewers expressed that a single player's ability to clear the scoreboard, as in many arcade games, could lead to higher engagement and an extra objective (especially targeted towards the players who seek completion as an accomplishment). The updated score system was received positively by the reviewers and were considered a large upgrade. Figure 11 shows how the new high score system is presented to the player through the game over screen with a leader board. It also illustrates the updated fonts and buttons in the game.

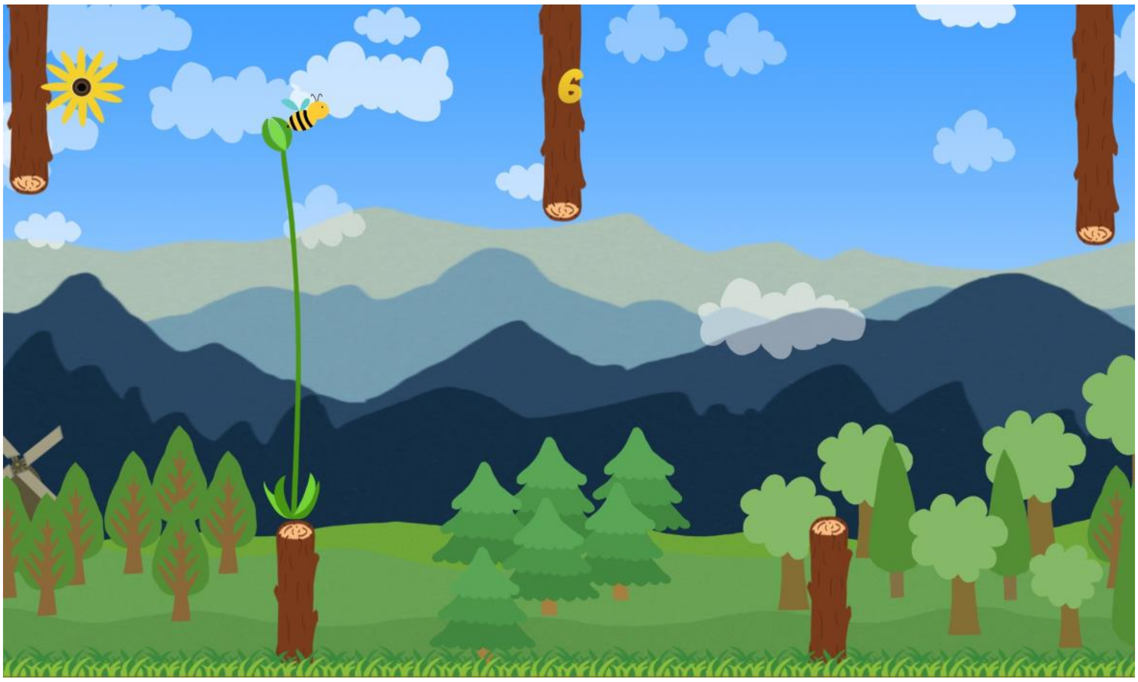


Figure 10 The updated main level in HeartBee.

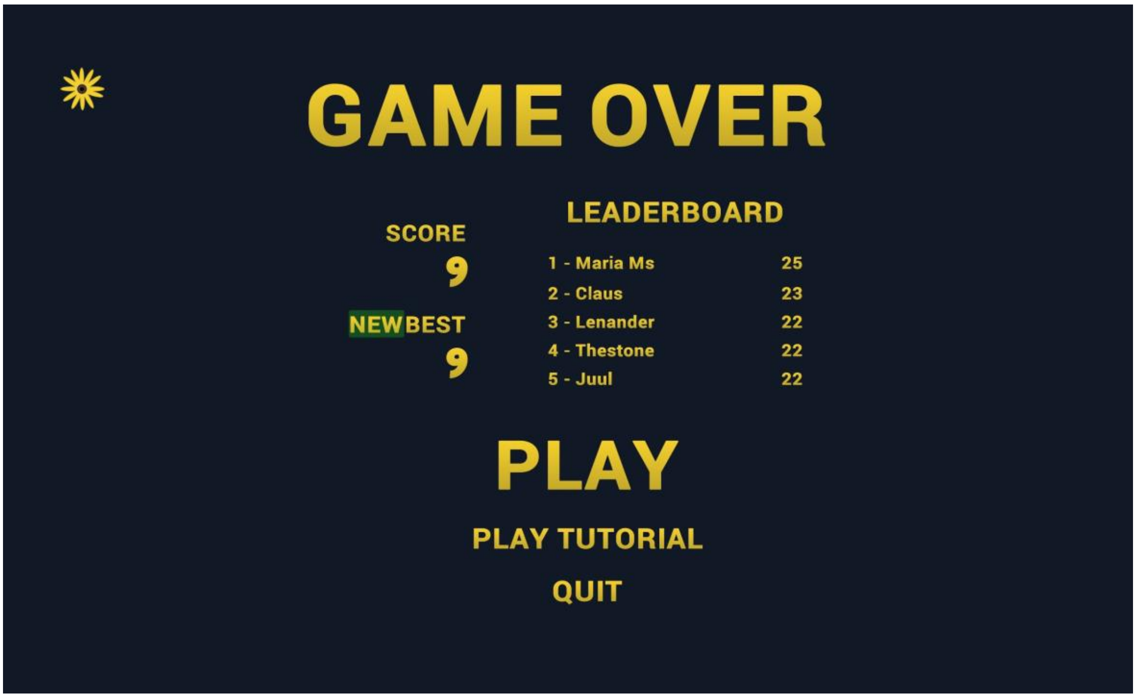


Figure 11 The updated game over screen in HeartBee.

General UI updates - Amongst general UI updates is the implementation of buttons to help the player navigate through the game. The pre-thesis game was controlled all automatically. In the current version, players are given the possibility to play the game again, play the tutorial again, or quit the game. This gives the player a larger agency over what happens.

In addition to buttons, a "lost connection" text was designed to pop up in the lower right corner whenever the sensor did not send readings. This was done to give users clear feedback, even though it is feedback about no feedback.

Finally, a progress bar was added to the initial stage of the game where the system is collecting initial data - again to improve the feedback during the whole experience.

Updates to player movement - As revealed in the first iteration, players like a mixture of the pre-thesis control scheme and the new control scheme. The final design merges the two so that the bee directly follows the player's HR during normal play whereas the bee uses "autopilot" to continue in the previous trajectory of the player when losing connection. In addition, the main movement was smoothed rather than following a constant speed.

Finally, observations showed that becoming invincible for collecting flowers led to a lost focus on breathing. This hurt the players more than it benefited them. To counter this, collecting a flower was designed to give an extra point rather than invincibility. It still provides a follower for the player.

In addition to the abovementioned reviewed updates, the game was finally built for the Android platform which required a re-scaling of most of the UI. This concludes the design iterations in relation to enhancing the player experience. In the following section I describe how the design of the game was tweaked in accordance to the experimental test.

4.4 The design in relation to the experiment

The pre-thesis game was designed as part of an experiment in which HRV was measured. Due to the literature stating that measurements should be exactly equal in length when comparing HRV, the game included two rest levels in which HRV was captured while the user was passive. For this thesis, such measurements are not necessary, and the rest levels have therefore been removed from the game.

The pre-thesis version was fixed to a 10-minute play session, after which the experiment would end, and the game would shut down. In this thesis I aim to test for how long players are willing to play the game. The maximum time limit set by the experiment is 15 minutes. The timer has therefore been changed accordingly.

Buttons, as explained in the previous section, have been added to the game to give the player agency over the experience and when it ends.

In relation to flow (Csikszentmihalyi, Flow: The psychology of optimal experience, 1990), I argue that the game addresses all nine requirements for achieving flow:

A challenging activity that requires skills - Despite it's simple nature - all you need to do is breathe - the game is a challenging activity that requires skills. The challenge level of the game is also based on the player's own physiology and as such the game tries to adapt the challenge to the skills of the player.

The merging of action and awareness - Players tend to become very aware of their breath during gameplay as that is the core mechanic and they do not notice how people might view them while doing so.

Clear goals - The game has clear extrinsic goals - get as far as possible, but also lends itself to intrinsic goals, e.g. bettering one's control over the HR.

Clear feedback - The feedback is clear in the sense that it is a direct visualization of the player's HR. The player can immediately see how respiration affects the HR.

Concentration on the task at hand - The game requires a high level of concentration on the task at hand. The control over the HR quickly disappears if the player stops concentrating on the breath.

The paradox of control - The player will quickly learn that the HR is easily controlled through breathing and the player will therefore know that the bee can be controlled through focusing on the task at hand.

The loss of self-consciousness - During play, the player can easily be so focused on the game that consciousness of the self and the surroundings is lost. This is especially evident when the player is about to set a new high score.

The transformation of time - The endless runner nature of the game makes it very easy to “just try one more time”. This can lead to the player playing the game for long periods of time without noticing that time has passed.

The autotelic experience - The purpose of HRV biofeedback training is that the player can reach a more relaxed state and overall well-being. As such, the player might discover a sense of relaxation through playing the game which could become the main objective for the player, leading to an autotelic experience.

In the following I analyse the game in relation to the OA3 framework (Schoenau-Fog, 2011b) to identify the conceptual categories of the framework that can be identified in the game as engagement triggers:

Objectives - The game lends itself to both extrinsic objectives and intrinsic objectives. An extrinsic objective can be to reach a certain score whereas an intrinsic objective can be to gain control over the HR.

Accomplishment - Completion can be a trigger for engagement through clearing the scoreboard as expressed by one of the reviewers. Progression was confirmed to be a possible trigger for engagement through getting a sense of control over the bee and through getting better scores over time.

Activities - Socializing, sensing, and interfacing have all been confirmed as engagement triggers through the pre-thesis experiment as well as through the iterative design process of this thesis. Socializing can trigger engagement through the competitive element of a high score. Sensing can trigger engagement through the positively reviewed audio-visual landscape. Interfacing can trigger engagement through the alternative control scheme.

Affect - The game lends itself nicely towards positive emotions through the audio-visual landscape. Furthermore, improving on one’s own score can trigger satisfaction. Relaxation was confirmed to be a positive affect through user testing. The game lends itself to absorption through e.g. flow and to some extent, immersion. Frustration is also a possible trigger for engagement, especially if a player has a hard time beating a score but refuses give up.

This concludes the design chapter of this thesis. The design of the game has been described through iterative user testing. How the game addresses the requirements of the experiment has also been elaborated upon. The following chapter describes how the design was carried out from an implementation perspective.

5 IMPLEMENTATION

In this chapter I present some of the key elements implemented to realize the final design. Specifically, the focus is on implementing the Empatica E4 wristband - both for standalone applications and for Android, and on implementing an online high score system.

First, however, I provide an overview of the general elements of the implementation.

5.1 Overview

HeartBee is developed in Unity (Unity Technologies, 2018) game engine (version 2018.2.1f1). The game was first developed as a standalone application for testing purposes and was later developed as an Android gradle project in Android Studio (Google, 2018). I borrowed a Samsung Galaxy S7 smartphone from the SMILE lab at AAU CPH for developing and testing the game. The screen resolution of the Galaxy S7 is 1920x1080. The game was developed with this resolution in mind.

The game code is written in C# using the community version of Visual Studio 2017 (Microsoft, 2018). The Empatica E4 code for android is written in Java.

The music used in the game is royalty free music created by Bensound (Bensound, 2018).

The Unity project contains four scenes: Setup, Tutorial, Main, and Final. Setup is the scene in which connection is established and the player chooses a player name. Tutorial is the scene where the player is taken through the tutorial of the game. Main is the gameplay scene. Final is the scene that contains information about player scores and the leader board.

Each scene contains a unique canvas object for handling UI and a camera for rendering the scene. The canvas is set to scale with screen size and has the 1920x1080 resolution as a reference resolution. In addition to this, the tutorial scene and the main scene both contain a player object, and the main scene contains objects for generating and updating the scene.

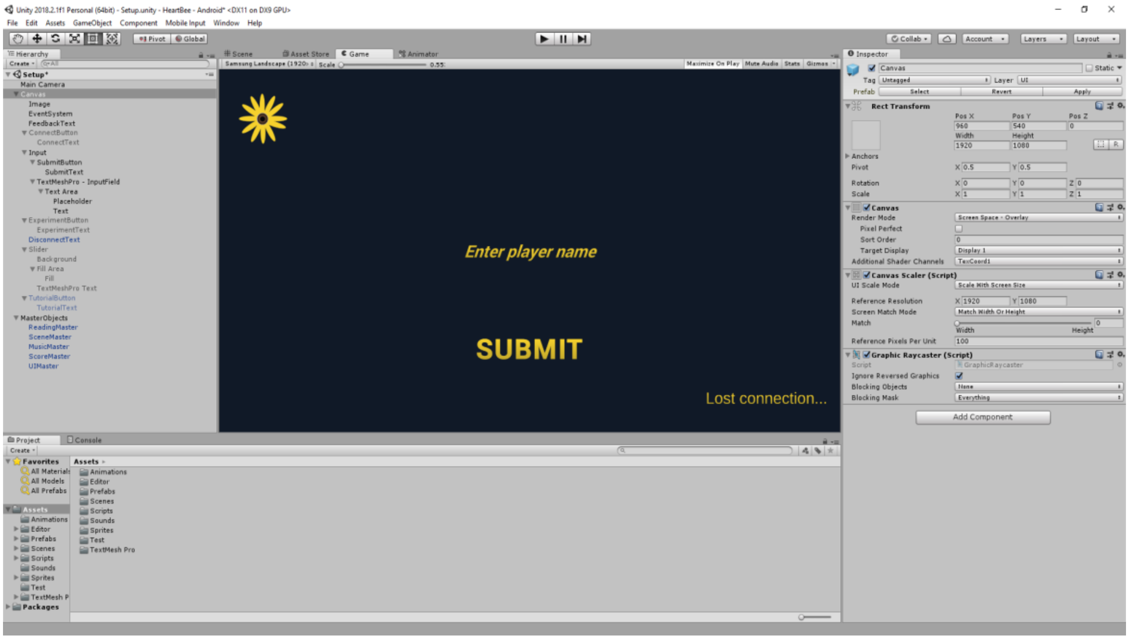


Figure 12 A screenshot of the Setup scene. The Hierarchy shown is common for all scenes.

The setup scene contains a set of “master” objects, all developed as singletons. A singleton means that the class can instantiate only one of such objects. This is useful when referencing it from other scripts as one does not have to find the object first.

These are the ReadingMaster object (handles and stores readings from the E4), the SceneMaster (handles loading scenes and sending messages to the other master objects once the scenes are loaded), the MusicMaster (handles playing music during gameplay), the ScoreMaster (handles the in-game score as well as the online high scores), and the UIMaster (handles UI related functions). The structure and layout of the Setup scene is shown in Figure 12.

This sums up the overview of the implementation. The following section describes how the E4 wristband was integrated into the game.

5.2 Integrating the Empatica E4 wristband

Integrating the Empatica E4 wristband into the game was an important step towards making the game mobile. The game was mainly developed on Windows. This was chosen as the only way to build it for Android was to build it as an Android gradle project which should then be merged with the Android code provided by Empatica. In addition to that, the manifest file of the Android project had to be modified. This meant that every change that was made in the game required a new Android build and a new merge, making it a very tedious task. Most tweaks to the game were thus made using the Windows connection and the project was only built for Android for testing purposes whenever a major milestone in the implementation had been reached and needed testing.

The E4 wristband is a wireless wristband, much like a smartwatch, that communicates using the Bluetooth low energy (BLE) communication protocol. The E4 collects a multitude of data: EDA, BVP, acceleration, HR, and temperature. The wristband is developed for use in clinical studies by researchers. The E4 collects HR data by analysing the BVP signal. The BVP signal is collected through PPG (illuminating the skin and measuring the reflected light) and is sampled at 64Hz. IBI are gathered through a peak detection algorithm applied to the BVP signal.

To use the E4 for personal applications, one must first sign up on the website as a developer to receive a unique API key that needs to be specified for development. In addition to that, the wristband must be linked to the API key by registering a purchase code. The E4 needs internet connection to connect to the application to be able to confirm that the product is allowed with the API key.

5.2.1 Communicating with E4 on Windows

For development on the windows platform, Empatica has released the E4 streaming server

```
public static void SetupStream(string deviceName)
{
    try
    {
        _client = new TcpClient(_hostName, _port);
        _stream = _client.GetStream();

        string response = "";
        WriteStream("device_connect " + deviceName);
        response = ReadStream();
        string[] words = response.Split(' ');
        if (!words[words.Length-1].Contains("OK"))
        {
            throw new ArgumentNullException(response);
        }
        WriteStream("device_subscribe ibi ON");
        response = ReadStream();
        if (!words[words.Length - 1].Contains("OK"))
        {
            throw new ArgumentNullException(response);
        }
        socketReady = true;
    }
    catch (ArgumentNullException e)
    {
        Debug.Log("ArgumentNullException: " + e);
    }
    catch (SocketException e)
    {
        Debug.Log("SocketException: " + e);
    }
}
```

Code Snippet 1 The SetupStream method of the StreamingThread class.

(Empatica, 2018). The streaming server is an application that can connect to the E4 device and stream the signal to other applications through a TCP communication protocol. The streaming server is only able to connect to an E4 with a BlueGiga BLED112 Bluetooth Smart Dongle (Silicon Laboratories, 2018). As such, the streaming server needs to run in the background for other applications to receive data from the E4 wristband.

Connecting to, communicating with, and handling the data from the E4 is done in the StreamingThread class of the Unity project. The StreamingThread class is a public static class, meaning that there is no instance instantiated in the project. Instead, all other classes can call it easily and get access to its data. Code Snippet 1**Error! Reference source not found.** shows the SetupStream method that uses the unique communication protocol of the E4 streaming server.

The first steps of the SetupStream method is to establish a TCP connection to the specified host and reference the data stream of this connection. The unique communication protocol then starts: First, the Unity application sends a message to the streaming server, asking to connect to the device. The message must be formatted as "device_connect devicename". The streaming server then sends a response. This is either a) "R device_connect_btle OK" or b) "R device_connect_btle ERR <reason>". With this in mind, I split the string based on the whitespace. Then I check if the last of those strings is equal to OK which means that the game is connected to the E4. The game then tries to subscribe to the IBI data stream from the device. Successfully subscribing to the data stream means that the game continuously receives IBI data from the E4.

```
public static void StartReadingThread()
{
    _myThread = new Thread(new ThreadStart(ConnectionStream));
    _myThread.Start();
}

private static void ConnectionStream()
{
    while (_myThread.IsAlive)
    {
        string response = ReadStream();
        if (response.Contains("Hr"))
            continue;

        response = response.Replace("\r\n", "");
        response = response.Replace(",", ".");

        AddNewReading(response);
    }
}
```

Code Snippet 2 The StartReadingThread and ConnectionStream methods of the StreamingThread class.

IBI data is sent whenever the device registers a new pulse reading. IBI is the interval between the last pulse and the current pulse. The frequency of the received IBIs is then based on the pulse of the player. A mean HR of 60 BPM therefore means that IBIs are received at 1Hz.

Reading a socket from a TCP stream blocks the thread from which it is called until data has been received. This means that, when continuously reading the IBI, the code will run once a second. This is not acceptable for the game and would make it completely unplayable. To counter this problem, reading from the socket is performed in a separate thread. This is done in the ConnectionStream method which is started from the StartReadingThread method, both shown in Code Snippet 2.

The ConnectionStream method is started by _myThread, which is a new Thread object. The stream is then running as long as the thread is alive through a while loop. In the loop, a response is first received from the data stream. When subscribing to the IBI data stream, the game receives both the IBI and the current HR (which is just the IBI turned into HR). Thus, I am only interested in the response that contains the IBI. This is handled by the if statement that continues to the next iteration (i.e. skips the remaining steps of the while loop to start from the beginning again).

```
public static void AddNewReading(string msg)
{
    var subStrings = msg.Split(' ');

    double time = double.Parse(subStrings[1]);
    float ibi = float.Parse(subStrings[2]);
    IBIReading reading = new IBIReading(time, ibi);
    lock (_lock)
    {
        unhandledReadings.Enqueue(reading);
    }
    readingsReceived++;
    EvaluateReading(reading);
}

public static IBIReading GetNewReading()
{
    lock (_lock)
    {
        IBIReading reading = unhandledReadings.Dequeue();
        handledReadings.Enqueue(reading);
        if (handledReadings.Count > 30)
            handledReadings.Dequeue();
        timeOfLastReading = Time.time;
        return reading;
    }
}
```

Code Snippet 3 The AddNewReading method and the GetNewReading methods of the StreamingThread class.

Through analysing the strings received, I noticed two things that had to be handled before storing the data - 1) the string response were terminated with the Windows newline command, “\r\n”, and 2) the data was formatted using commas instead of dots for floats and doubles. The newline character is therefore removed, and the commas are replaced with dots before handling the response. Handling of the response is done through the AddNewReading method which is shown together with the GetNewReading method in Code Snippet 3.

The AddNewReading method takes the formatted message as input and splits it into substrings based on the whitespace. The formatted message is in the format of “E4_Ibi time IBI”. The first part of the string is redundant information for processing HRV and it is therefore not used. The second part, time, is Unix time, and thus needs the formatting of double. The second part, IBI, is the interval in seconds since the last heartbeat. A new object is then created of the type IBIReading. IBIReading is a small class that was created to keep the time and IBI of each measurement together. The reading is then added to a queue of unhandled readings (readings that have not been processed by the main thread yet). Due to two separate threads having access to that queue, it is put behind a lock which means that it cannot be accessed by one thread if it is already being altered by another thread.

A counter for how many readings have been received is incremented and the reading is evaluated by updating the max, min, and mean BPM accordingly.

The GetNewReading method is called from the main thread. It removes the first IBI reading “in line” and adds it to a queue of readings that have been handled. Handled readings is essentially a list of the last 30 readings which is used to evaluate the max, min, and mean BPM.

Finally, I keep track of the time this reading was provided. This is used to check if there is a weak signal by keeping track of if the time passed since a reading is larger than expected, i.e. 2 seconds. If so, the “Lost connection” text, as explained in the design chapter, becomes visible on screen.

This sums up how the game communicated with the E4 on a Windows computer. In the following section, I explain how the communication is handled from an Android device.

5.3 Communicating with the E4 on Android

Whereas building the game for android and running it was easy, getting it to connect with the E4 proved to be a big challenge due to various setup that had to be performed. First, the project requires the Empalink SDK to be downloaded and added to the libs folder of the project.

Secondly, the E4 requires a connection to the internet, a Bluetooth connection, and access to the

```
<uses-permission android:name="android.permission.INTERNET" />
<uses-permission android:name="android.permission.ACCESS_COARSE_LOCATION"/>
<uses-permission android:name="android.permission.BLUETOOTH" />
<uses-permission android:name="android.permission.BLUETOOTH_ADMIN" />
```

Code Snippet 4 The permissions needed in order to use the E4 device

location of the device. This was added by adding the lines of Code Snippet 4 to the Android Manifest of the project.

Thirdly, dependencies on the Okhttp 2.5 and the Empalink SDK had to be added to the build.gradle file. This is done by changing the dependencies method of build.gradle as in Code Snippet 5.

The above steps allowed the android application to connect to the E4 device. To do so, the Empalink sample project (Empatica, 2018) provided by Empatica was used as a guidance.

The Android project contains only one Activity. An Activity is essentially the same as a scene in Unity, but since the scenes had already been set up in one activity, there was no need to implement more. To communicate with the E4, the main activity had to implement the EmpaStatusDelegate and the EmpaDataDelegate interfaces and their methods. These methods include listener methods for receiving the various datastreams from the E4.

An extensive implementation of the connection can be found in both the Empalink sample project and in the HeartBee Android project in the digital appendix. What is unique in the game, however, is the need to communicate with Unity. The protocol for doing so is shown in Code Snippet 6, the didReceiveIBI method.

The didReceiveIBI method is called whenever the E4 registers a reading. As such, this means that there is no need for multithreading in Unity as there is no reading data from a stream. The message is then converted into the same format as from the streaming server.

Finally, the method calls the UnityPlayer.UnitySendMessage method. This is the method that lets Android communicate with a Unity project. To pass the message, the method takes

```
dependencies
{
    implementation fileTree(dir: 'libs', include: ['*.jar'])
    implementation 'com.empatica.empalink:empalink:2.2@aar'
    implementation 'com.android.support:appcompat-v7:27.1.1'
    implementation 'com.squareup.okhttp:okhttp:2.5.0'
}
```

Code Snippet 5 Dependencies needed to use the E4 device.

```
public void didReceiveIBI(float ibi, double timestamp) {
    String msg = "ibi " + timestamp + " " + ibi;
    Log.d("unity", msg);
    UnityPlayer.UnitySendMessage("ReadingMaster","AddReadingFromAndroid",
    msg);
    DoSomething();
}
```

Code Snippet 6 The didReceiveIBI listener method of the HeartBee android project.

in the name of the game object that it is trying to communicate with (i.e. ReadingMaster), the name of the function that it calls on that object (i.e. AddReadingFromAndroid), as well as the parameter of the function (must be a string). From a Unity perspective, the AddReadingFromAndroid function of the ReadingMaster singleton simply calls the AddNewReading method of the StreamingThread class.

This sums up the description of how the E4 communicates with HeartBee through both a windows application and an Android application. The steps for setting up the android project (update manifest, update build.gradle, and merging with the Empalink sample project) had to be done whenever a new Android project was built, i.e. whenever there was a change in the game. Therefore, the Windows version was beneficial during prototyping.

5.4 Implementing a high score system

It was established through the design chapter that there was a high demand for a high score system and a leader board in the game. The pre-thesis game stored the best score achieved locally through Player Preferences. Expanding the Player Preferences to store the five best scores instead of just one would be an easy and quick solution for implementing a leader board. This solution, however, is limited to one device, and players would therefore not be able to compare scores across devices, something that would be required if the game was ever released. The solution to this problem was to implement an online high score system. Normally, developing such system would require an online database. To make it easier for developers, however, Carmine T. Guida developed Dreamlo (Guida), a free online score system that works on several platforms, including Unity. By default, Dreamlo stores up to 1000 of the best scores.

Dreamlo generates a personal private code and public code that can be used to upload and fetch scores from a script respectively. A limitation to the system, however, is that it only stores one score per name. This generates two main problems - 1) a player cannot clear the scoreboard, as was requested in the design, and 2) two players with the same name will overwrite each other’s scores. The first of these problems were solved by adding an index (from 0 to 4) to the end of the name of the player for each score they got. The second problem is yet to be solved but was

```
IEnumerator DownloadHighscoresFromDatabase(bool limitToTop)
{
    float time = Time.time;
    string myUrl = webURL + publicCode + "/pipe/";
    if (limitToTop)
        myUrl += highscoreLimit;
    WWW www = new WWW(myUrl);
    yield return www;
    if (string.IsNullOrEmpty(www.error))
        FormatScores(www.text);
    else
        print("Error downloading: " + www.error + " -- Time before timeout: "
+
                (Time.time-time));
    if (!limitToTop)
        LoadLocallyStoredScores();
}
```

Code Snippet 7 The DownloadHighscoresFromDatabase coroutine of the ScoreMaster script.

accounted for during the experiment by making sure that the names chosen by players were not already in use.

The score system, found in the ScoreMaster script, is a relatively extensive implementation with more than 300 lines of code. A few snippets have thus been chosen that shows the general workflow. The full code can be found in the digital appendix.

Downloading and uploading scores to the high score system is done using a WWW object, which allows simple access to web pages. Code Snippet 7 shows the DownloadHighscoresFromDatabase coroutine which is the implementation of how scores are downloaded.

The DownloadHighscoresFromDatabase coroutine takes a Boolean, limitToTop, as its parameter. It is possible to limit the number of scores downloaded from the website. Initially, when the game opens I want to download all scores. This will also fetch the players' own best score if they made it to the top 1000. Afterwards, however, there is no need to download more scores than what can be displayed in the leader boards and I therefore limit this.

A WWW object is then created that tries to fetch the scores from the webpage. The returned object contains information about errors, and if there is none, I format the downloaded scores and add them to the game. Additionally, the system loads locally stored scores (scores that were unable to be uploaded last time the game was played) when the game starts.

This coroutine runs continuously through the game in intervals of 30 seconds to keep the leader boards relatively up to date.

Uploading scores to Dreamlo can either be done alone or together with downloading scores. In most cases, downloading the scores while uploading is preferable as it saves a trip to the database. The method for coroutine for doing so, UploadAndFetchScores, is shown in Code Snippet 8.

The UploadAndFetchScores coroutine takes a ScoreInstance object as its input. The ScoreInstance class is a small class for storing information about scores (name, score, index) together. The name is then formatted as “name,index” and uploaded to the database with the WWW object. Like with the previous coroutine, this coroutine then waits for action to complete, and formats the downloaded high scores.

This sums up the implementation chapter of this thesis. As explained previously, this is only a small selection of the complete implementation which can be found in the digital appendix with commented code. In addition to the presented implementation, the game also stores the names, scores, time spent, and IBI data in csv files after each playthrough. This can be used for later evaluation. In the following chapter, I present the procedure and results of the experiment that was carried out.

```
IEnumerator UploadAndFetchScores(ScoreInstance scoreInstance)
{
    string userName = scoreInstance.name + "," + scoreInstance.idx;
    int score = scoreInstance.score;
    WWW www = new WWW(webURL + privateCode + "/add-pipe/" +
                  WWW.EscapeURL(userName) + "/" + score + "/");
    yield return www;
    if (string.IsNullOrEmpty(www.error))
    {
        FormatScores(www.text);
        scoreAddedToDB = true;
    }
    else
        print("Error uploading: " + www.error);
}
```

Code Snippet 8 The UploadAndFetchScores coroutine of the ScoreMaster script.

6 TEST METHOD

In this chapter I describe the participants and procedure of the final experiment. The test is based on the experimental design presented in the Methodology chapter.

6.1 Participants

The participants recruited for the final experiment were gathered through convenience-sampling to reach as many participants as possible. Participation in the experiment was voluntary. It was required that all recruited participants were between 18 and 44 years old. Participants were required to either have an undergraduate degree or be studying at a university. Alternatively, they had to be users of mHealth applications. These requirements were put forth to only recruit people who were more likely to adapt mHealth applications themselves, as explained by (Carroll, et al., 2017).

Initially, 32 participants were recruited to take part in the experiment. One participant was unable to play the game due to the sensor not being able to register his pulse consistently. He was excluded from the final sample.

The final sample ($N = 31$) consisted of 11 students. The remaining were either employed or between jobs but had all, except for one, completed an undergraduate education. The one who did not was an avid user of mHealth applications. The sample consisted of nine females. The average age of the participants was 27 ± 3.7 and ranged from 21-39. 29 participants were Danish. The remaining two were Dutch.

Random allocation of the participants put 16 participants in the HRCP first group. The remaining 15 were allocated to the HeartBee first group.

6.2 Procedure

The experiment was conducted in Copenhagen from the 14th to the 21st of August 2018 in a time frame spanning from 10:00 am to 10:00 pm. Experiments were carried out at locations convenient for the participants - three participants were tested at AAU CPH. Another three participants were tested in the home of the researcher. The remaining 25 were tested in the comfort of their own home. It was made sure that the test could be carried out in a controlled environment before a location was accepted. A controlled environment meant that the participants would not be disturbed during test.

Before starting the experiment, participants were informed that the experiment would last for up to 45 minutes. They were informed that they would be trying out two different Android applications for measuring and controlling their HR during breathing exercises. They were not informed about who developed which application until after the experiment had been completed. The participants were then asked to sign a written consent and fill out their demographics (gender, age, nationality, and occupation) on the online questionnaire.

The participants were introduced to the first application after filling out the demographics. The first application chosen was random - 16 of the participants started with HRCP while the remaining 15 started with HeartBee.

The participants who started with HRCP were given a brief introduction to the UI and the measurements of the application and were informed to start the experience whenever they wanted by pressing the heart on the screen.

The participants who started with HeartBee were asked to open the application and put in their player name. They were informed that the player name might be referred to in this thesis and they were thus encouraged to use an alias if they wanted to stay anonymous. They were then introduced to how the E4 wristband worked and how to reduce noise from the sensor (i.e. sit still during the experiment) with the purpose of getting a better reading. They were then asked to turn on the E4 wristband and start the experience. All other guidance was provided by the game.

For both applications, the participants were informed that they should stop the experience at any time they wished. They were also informed that the researcher would not gain more data by using the application for longer. The participants were informed to let the test conductor know when they were done using the application.

The participants were asked to open the questionnaire again after stopping their experience and fill out the form which contained the questions related to flow and continuation desire. They were allowed to ask the test conductor if they had a hard time understanding the questions.

Figure 13 A participant playing HeartBee during the experiment.

After filling out the form, the participants were introduced to the other application (HeartBee or HRCP, depending on which they started with). The participants who got HeartBee second followed the same procedure as the ones who started with HeartBee. The participants who got HRCP second followed the same procedure as the ones who started with HRCP.

Finally, after finishing the second application, participants were asked to fill out the last part of the questionnaire which consisted of the same items as the first one.

When done, participants were thanked for taking part in the experiment and were debriefed about the purpose of the experiment. Figure 13 shows a test participant playing HeartBee as part of the experiment.

The experiment was carried out on one to two participants at a time. The devices used for carrying out the experiment was a Samsung Galaxy S7, borrowed from the SMILE lab at AAU CPH, and a Huawei Mediapad M5 10.8.

This concludes the test method from a participants and procedure perspective. The following chapter presents the evaluation of the experiment.

7 EVALUATION

This chapter deals with presenting the findings of the final experiment. The findings presented are divided into three sections - The first section deals with evaluating the three main hypotheses of the thesis against the corresponding null hypotheses. The second section deals with the three secondary hypotheses. The third section is a presentation of the qualitative data gathered from the continuation desire framework.

7.1 Evaluating the main hypotheses

This section deals with reporting the findings from testing the three main hypotheses of the experiment:

- **H1**: The HRV biofeedback game elicits higher levels of flow than the non-game HRV biofeedback application.
- **H2**: The HRV biofeedback game elicits higher levels of continuation desire than the non-game HRV biofeedback application.
- **H3**: The HRV biofeedback game leads to more time training than the non-game HRV biofeedback application.

The hypotheses are tested against the null hypothesis that there is no difference in the measures of engagement between HeartBee and HRCP. Significance is accepted at a 95% threshold for confidence, i.e. $p < 0.05$.

The chosen experimental design for evaluating the main hypotheses was a counterbalanced repeated-measures design. This calls for a Two-Way Mixed ANOVA. This test splits the group into two factors - a within-group factor and a between-group factor. The within-group factor is the treatments (i.e. game or non-game) and the between-group factor is the order of presentation. For readability purposes, the group that was presented with HeartBee first is referred to as Group A. The group that was presented with HRCP first is referred to as Group B. The Two-Way Mixed ANOVA assumes homogeneity of variance. This assumption holds true for all reports unless specified otherwise in the individual evaluations.

The Two-Way Mixed ANOVA are followed up with post hoc tests. Specifically, a Paired Samples T-test is used for parametric data and a Wilcoxon Signed Rank test for non-parametric data.

7.1.1 Flow

Flow was measured two times for all participants - after playing HeartBee and after using HRCP. Flow was measured using the SFSS with statements such as "*I found the experience extremely rewarding*". Possible scores ranged from 1 (strongly disagree) to 5 (strongly agree).

Both measures of flow fulfilled the requirements of parametric statistics and are therefore followed up by a Paired Samples T-test. Table 4 shows the mean and standard deviation (SD) of the flow scores for both groups during the two treatment conditions (HeartBee and HRCP).

A Two-Way Mixed ANOVA was conducted on the flow scores. The results revealed that the flow scores were significantly different between the two treatment conditions, $F(1,29) = 4.81, p = 0.04, r = 0.14$. Specifically, participants reported higher levels of flow during HeartBee ($M = 3.85$) than during HRCP ($M = 3.56$), $t(30) = 2.22, p = 0.04, r = 0.38$.

The results revealed no significance of group, either as a main effect ($F(1,29) = 0.61, p = 0.44, r = 0.02$), or as an interaction with the treatment condition ($F(1,29) = 1.25, p = 0.27, r = 0.04$). Thus, participants reported higher flow levels after playing HeartBee than they did after using HRCP. This was true regardless of whether they tried HeartBee or HRCP first.

Table 4 Flow scores as a function of group and treatment version.

Group	HeartBee		HRCP	
	Mean	SD	Mean	SD
Group A	3.84	0.43	3.70	0.75
Group B	3.86	0.60	3.43	3.39

Table 5 Continuation desire scores as a function of group and treatment version.

Group	HeartBee		HRCP	
	Median	IQR	Median	IQR
Group A	5.00	2.00	3.00	2.00
Group B	6.00	2.00	4.50	2.00

7.1.2 Continuation desire

Continuation desire was measured two times for all participants - after playing HeartBee and after using HRCP. Continuation desire was measured using the post-game question of the basic game ESQ, i.e. "*I want to try again*". Possible scores ranged from 1 (strongly disagree) to 7 (strongly agree).

None of the measures of continuation desire fulfilled the requirements of parametric statistics and are therefore followed up by a Wilcoxon Signed Rank test. Table 5 shows the median and interquartile range (IQR) of the continuation desire scores for both groups during the two treatment conditions (HeartBee and HRCP).

A Two-Way Mixed ANOVA was conducted on the continuation desire scores. The results revealed that the continuation desire scores were significantly different between the two treatment conditions, $F(1{,}29) = 31.23, p < 0.001, r = 0.52$. Specifically, participants reported higher levels of continuation desire during HeartBee than during HRCP, $T = 15, p < 0.001, r = 0.51$.

The results revealed no significance of group, either as a main effect ($F(1{,}29) = 2.09, p = 0.16, r = 0.07$), or as an interaction with the treatment condition ($F(1{,}29) = 0.15, p = 0.7, r = 0.01$). Thus, participants reported a higher desire to try HeartBee again than trying HRCP again. This was true regardless of whether they tried HeartBee or HRCP first.

7.1.3 Time spent

The time that participants spent on the different treatments was measured. The measurement reported is in seconds. Sessions were stopped after 900 seconds (15 minutes) and the time spent is therefore capped at this number. Possible scores thus range from 0 to 900.

The time spent during HeartBee was significantly skewed and could therefore not be considered as parametric. The follow up tests are thus performed by a Wilcoxon Signed Rank test. Table 6 shows the median and interquartile range (IQR) of time spent for both groups during the two

Table 6 Time spent in seconds as a function of group and treatment version.

Group	HeartBee		HRCP	
	Median	IQR	Median	IQR
Group A	900.00	165.71	213.00	199.00
Group B	888.54	99.36	281.50	139.50

treatment conditions (HeartBee and HRCP). It appears that, regardless of group, participants spent more time playing HeartBee than they did using HRCP.

A Two-Way Mixed ANOVA was conducted on the time spent. The results confirmed the initial impression: time spent was significantly different between the two treatment conditions, $F(1,29) = 498.8, p < 0.001, r = 0.95$. Specifically, participants spent more time playing HeartBee ($Mdn = 900$) than they did using HRCP ($Mdn = 271$), $T = 0, p < 0.001, r = 0.68$.

The results revealed no significance of group, either as a main effect ($F(1,29) = 0.78, p = 0.36, r = 0.03$), or as an interaction with the treatment condition ($F(1,29) = 0.39, p = 0.59, r = 0.01$). Thus, participants spent longer time playing HeartBee than they did using HRCP. This was true regardless of whether they tried HeartBee or HRCP first.

The three main hypotheses of the thesis were thus confirmed: Participants rated their level of flow and their desire to try again significantly higher during HeartBee than during HRCP. They also spent significantly longer time playing HeartBee than they did using HRCP.

7.2 Evaluating the secondary hypotheses

This section deals with reporting the findings from testing the three secondary hypotheses of the experiment:

- **H4**: Higher levels of flow are related to higher levels of continuation desire.
- **H5**: Higher levels of flow are related to longer time spent training.
- **H6**: Higher levels of continuation desire are related to longer time spent training.

The hypotheses are tested against the null hypothesis that there is no relation between the measurements of engagement.

Describing their relationship can be used to validate the findings and triangulate the results of the main hypotheses. Two of the measures proved to be non-parametric, and as such, all correlations are performed using Spearman's rho. The correlations are presented from four different perspectives: One is the correlation between the measures during gameplay, one is the correlation during the usage of HRVB, one is a combination of the two, e.g. all measures of flow are combined into one vector, and one is the correlation between the changes in flow (Δflow) and continuation desire (Δcontinuation desire) from HRCP to HeartBee.

7.2.1 Relationship during HeartBee

A Spearman's correlation test was performed to determine whether the flow scores reported by the participants after playing HeartBee were related to the continuation desire scores of the participants after playing HeartBee. The results indicated that there was a significant moderate positive correlation between the two measures, $r = 0.42, p = 0.02$: Participants who reported higher levels of flow while playing HeartBee were more likely to want to try again.

A Spearman's correlation test was also performed to determine whether the flow scores reported by the participants after playing HeartBee were related to the time spent playing HeartBee. The results indicated that there was a significant moderate positive correlation between the two measures, $r = 0.41, p = 0.02$: Participants who reported higher levels of flow while playing HeartBee were likely to spend more time on it.

A final Spearman's correlation test was performed to determine whether the continuation desire scores reported by the participants after playing HeartBee were related to the time spent playing HeartBee. The results indicated that there was a significant moderate positive correlation

Table 7 Correlation between the three measurements of player engagement while playing HeartBee.

	Flow	CD	Time spent
Flow	1.00		
CD	0.42*	1.00	
Time spent	0.41*	0.50**	1.00

* means that the correlation is significant at the 0.05 level.

** means that the correlation is significant at the 0.01 level.

between the two measures, $r = 0.5, p < 0.01$: Participants who spend more time playing HeartBee were more likely to want to try again.

The results of the three correlations are summarized in Table 7.

7.2.2 Relationship during HRCP

A Spearman’s correlation test was performed to determine whether the flow scores reported by the participants after using HRCP were related to the continuation desire scores of the participants after using HRCP. The results indicated that there was a significant moderate positive correlation between the two measures, $r = 0.54, p < 0.01$: Participants who reported higher levels of flow while using HRCP were more likely to want to try again.

A Spearman’s correlation test was also performed to determine whether the flow scores reported by the participants after using HRCP were related to the time spent using HRCP. The results indicated that there was no significant correlation between the two measures, $r = -0.1, p = 0.6$: Participants who reported higher levels of flow while using HRCP were no more or less likely to spend more time using HRCP.

A final Spearman’s correlation test was performed to determine whether the continuation desire scores reported by the participants after using HRCP were related to the time spent using HRCP. The results indicated that there was no significant correlation between the two measures, $r = -0.13, p = 0.47$: Participants who spent more time using HRCP were no more or less like to want to try again.

The results of the three correlations are summarized in Table 8.

Table 8 Correlation between the three measurements of player engagement while using HRCP.

	Flow	CD	Time spent
Flow	1.00		
CD	0.54**	1.00	
Time spent	-0.10	-0.13	1.00

* means that the correlation is significant at the 0.05 level.

** means that the correlation is significant at the 0.01 level.

7.2.3 Relationship combined

The combined relationship is based on a combination of the two previous correlations. The flow measures, continuation desire measures, and time spent measures of HeartBee and HRCP are combined into three vectors with 62 observations.

A Spearman's correlation test was performed to determine whether the flow scores reported by the participants in any activity were related to their continuation desire scores in that activity. The results indicated that there was a significant moderate positive correlation between the two measures, $r = 0.54, p < 0.001$: Participants who reported higher levels of flow in general were more likely to want to try the activity again.

A Spearman's correlation test was also performed to determine whether the flow scores reported by the participants in any activity were related to the time they spent on that activity. The results indicated that there was a significant weak positive correlation between the two measures, $r = 0.27, p = 0.03$: Participants who reported higher levels of flow in an activity were more likely to spend time on that activity.

A final Spearman's correlation test was performed to determine whether the continuation desire scores reported by the participants in any activity were related to the time they spent on that activity. The results indicated that there was a significant moderate positive correlation between the two measures, $r = 0.54, p < 0.001$: Participants who spent more time on an activity were more likely to want to try that activity again.

The results of the three correlations are summarized in Table 9.

Table 9 Correlation between the three measurements of player engagement in any activity.

	Flow	CD	Time spent
Flow	1.00		
CD	0.54**	1.00	
Time spent	0.27*	0.54**	1.00

* means that the correlation is significant at the 0.05 level.

** means that the correlation is significant at the 0.01 level.

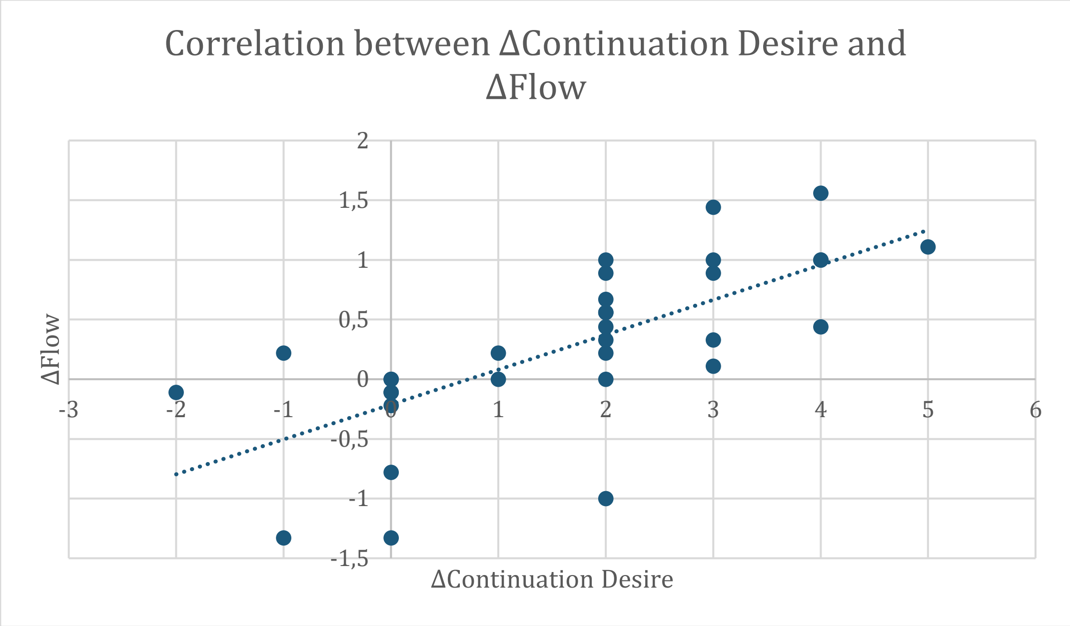


Figure 14 Correlation between Δflow and Δcontinuation desire.

7.2.4 Relationship between Δflow and Δcontinuation desire

The Δ relationship is based on the differences recorded between HeartBee and HRCP. The Δ relationship is considered because there might be differences in subjective baselines, e.g. one participant is more likely to rate a high desire to continue than another because of the individual idea of a desire to continue. To counter this problem, the Δ values are introduced. The Δ measurements consider one application (HRCP) as a baseline. As such, the scores are calculated by subtracting the scores of HRCP from the scores of HeartBee. This provides 31 values for Δcontinuation desire and 31 values for Δflow, one for each participant.

A Spearman's correlation test was performed to determine whether Δflow scores reported by the participants were related to their Δcontinuation desire scores. The results indicated that there was a significant large positive correlation between the two measures, $r = 0.7, p < 0.001$: Participants who reported a higher change in flow were more likely to report a higher change in continuation desire. Figure 14 shows the relationship between Δflow and Δcontinuation desire.

7.3 Evaluation of the game

The evaluation of the game is divided into two parts - the first part evaluates the flow components isolated and the second part deals with coding and evaluating the open-ended responses of the basic game ESQ reported after HeartBee and HRCP respectively. Evaluation of HeartBee and HRCP can be triangulated with the results of the main hypotheses and explain the findings in greater detail. Evaluating HeartBee can also lead to new design requirements for further development.

7.3.1 Evaluation of the nine flow components

Flow was measured using the SFSS. The SFSS aims to measure the nine components of flow. When isolated, the components (Likert items in the SFSS) are considered ordinal data. As such, the data is considered non-parametric. Table 10 shows the median and interquartile range (IQR) of the nine dimensions of flow for the two treatment conditions (HeartBee and HRCP).

A Friedman's ANOVA was conducted on the flow components for HeartBee. The results revealed that the flow components were scored significantly different, $\chi 2(8) = 81.55, p < 0.001$. Following up this finding with independent tests is inefficient due to the number possible combinations of variables. Instead, the components are inspected using the boxplot shown in Figure 15.

Table 10 Flow scores as a function of flow component and treatment version.

	HeartBee		HRCP	
	Median	IQR	Median	IQR
Challenging activity that requires skills	4.00	1.00	4.00	1.00
Merging of action and awareness	3.00	2.00	3.00	2.00
Clear goals	5.00	1.00	4.00	2.00
Clear feedback	4.00	1.00	3.00	2.00
Concentration on the task at hand	5.00	1.00	4.00	1.00
Paradox of control	3.00	2.00	4.00	1.00
Loss of self-consciousness	5.00	2.00	5.00	1.00
Transformation of time	4.00	1.00	3.00	2.00
Autotelic experience	4.00	1.00	3.00	2.00

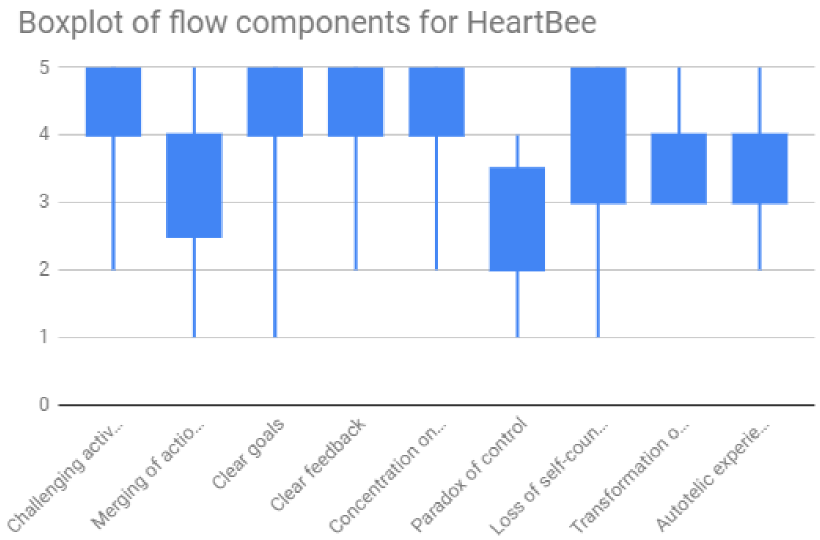


Figure 15 Boxplot of the scores of the nine flow components for HeartBee.

It appears from the boxplot that the components *challenging activity that requires skill*, *clear goals*, *clear feedback*, and *concentration on the task at hand* scored higher than the remaining components. Specifically, it appears that they were scored considerably higher than the *merging of action and awareness* and the *paradox of control* components.

A Friedman's ANOVA was also conducted on the flow components for HRCP. The results revealed that the flow components were scored significantly different, $\chi 2(8) = 71.11, p < 0.001$. Following up this finding with independent tests is inefficient due to the number possible combinations of variables. Instead, the components are inspected using the boxplot shown in Figure 16.

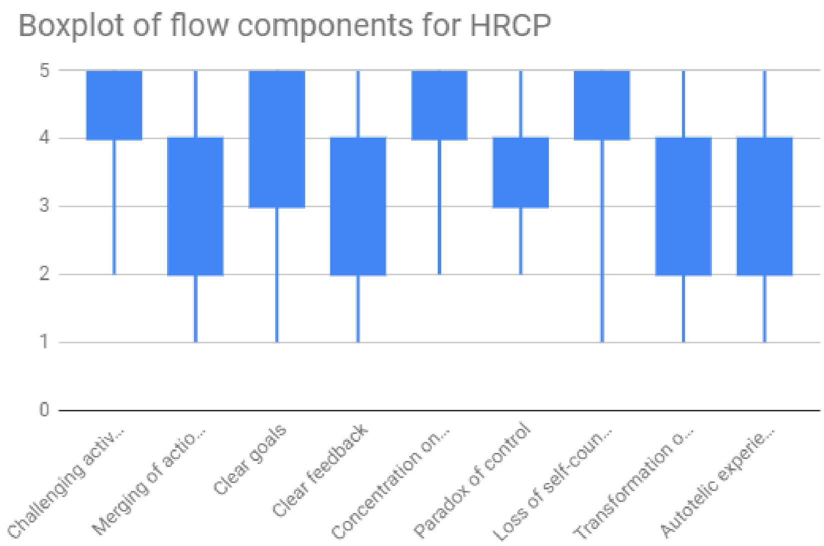


Figure 16 Boxplot of the scores of the nine flow components for HeartBee.

It appears from the boxplot that the components *challenging activity that requires skills*, *concentration on the task at hand*, and *loss of self-consciousness* were scored considerably higher than the remaining components.

The two boxplots suggest that participants had a stronger sense of clear goals and clear feedback while playing HeartBee than they did using HRCP. At the same time, participants felt a stronger sense of control while using HRCP than they did while playing HeartBee.

7.3.2 Qualitative analysis of HeartBee

The 31 qualitative responses related to HeartBee were analysed using traditional coding (Bjørner, 2015). Examples of responses are presented in the style of (Player name, gender, age). Traditional coding was used to organize, recognize, and code the responses. Specifically, each response was coded and evaluated against the 18 conceptual categories of the OA3 framework. The purpose was on finding engagement and disengagement triggers of HeartBee. The 31 participants who played HeartBee provided 45 statements related to engagement and disengagement in HeartBee: 36 were identified as engagement triggers and nine were identified as disengagement triggers. The identified engagement triggers for HeartBee are shown in Figure 17.

The most common engagement triggers were progression (e.g. "*I would like to improve my score!*" (Device, M, 26), positive affect (e.g. "*Because it was very relaxing and I was good at it*" (Maria Ms, F, 21), and interfacing (e.g. "*It was a fun new gaming experience that requires some new skills*" (Milou, F, 27)).

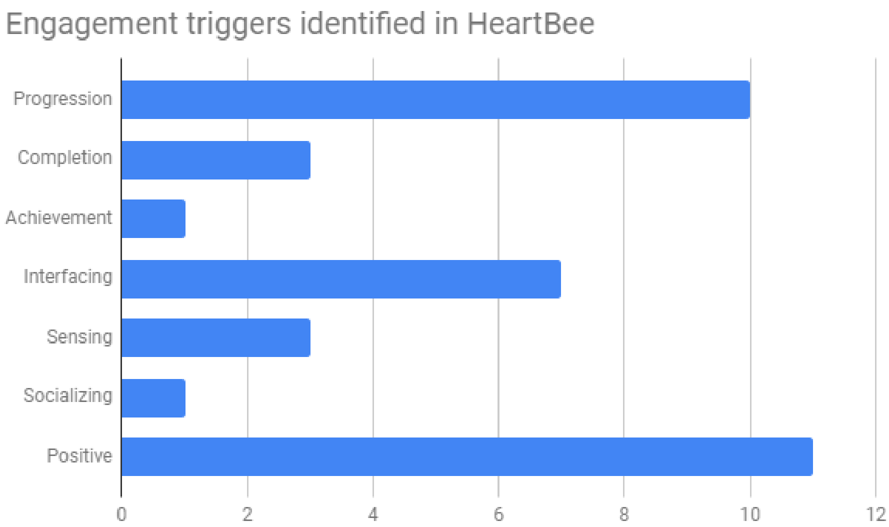


Figure 17 The identified engagement triggers for HeartBee.

The nine statements that identified disengagement triggers were related to interfacing ($N = 7$), negative affect ($N = 1$), and a lack of objectives ($N = 1$). Most of the statements regarding interfacing as a disengagement trigger ($N = 4$) were related to issues with the sensor, *e.g. "Was difficult but interesting, unfortunately the wristband lost connection every time*" (Jabba, M, 26).

7.3.3 Qualitative analysis of HRCP

The 31 qualitative responses related to HRCP were analysed using traditional coding. Examples of responses are presented in the style of (Player name, gender, age). Specifically, each response was evaluated against the 18 conceptual categories of the OA3 framework. The purpose was on finding engagement and disengagement triggers of HRCP which could lead to design implications for future HRV biofeedback applications and/or games.

The 31 participants who used HRCP provided 38 statements related to engagement and disengagement in HRCP: 16 were identified as engagement triggers and 22 were identified as disengagement triggers. The identified engagement triggers for HRCP are shown in Figure 18.

The most commonly reported engagement trigger was progression (e.g. "*It could be fun to see my initial result then try again to see if I could better my result*" (Jman, M, 26)), followed by positive affect (e.g. "*It was therapeutic and relaxing*" (Device, 26, M)), and interfacing (e.g. "*I liked the application as it was different from other typical mobile games and that I got to experiment with controlling my breath and heartrate*" (Ninjagirl, 25, F)).

The 22 statements that identified disengagement triggers were related to negative affect ($N = 13$), a lack of objectives ($N = 4$), a lack of progression ($N = 3$), bad sensing ($N = 1$), and bad interfacing ($N = 1$).

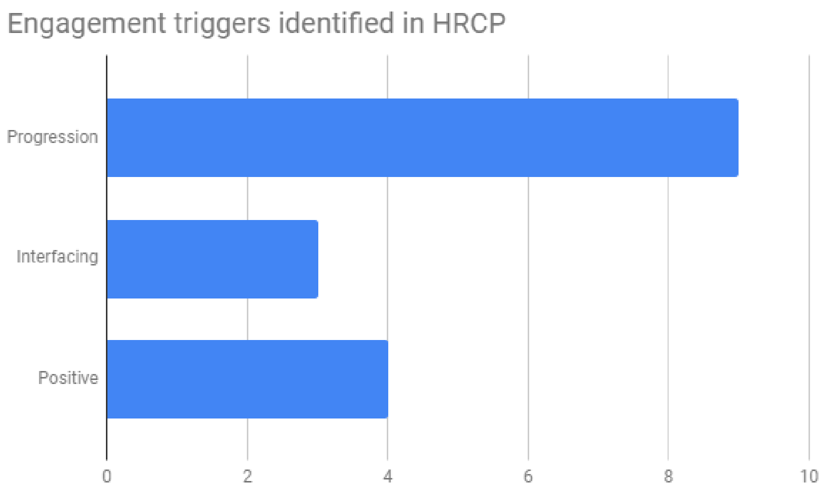


Figure 18 The identified engagement triggers for HRCP.

An analysis of the qualitative data from HeartBee and HRCP thus suggests that HeartBee lends itself to more engagement triggers than HRCP. Positive affect especially was more commonly reported during HeartBee. In comparison, negative affect was the most commonly reported disengagement trigger for HRCP whereas interfacing, mainly due to technical problems with the sensor, was the most commonly reported disengagement trigger for HeartBee.

This sums up the evaluation chapter. In the following chapter, the discussion, I analyse the findings and relate them to theory and experimental findings presented in the background chapter. The limitations of the experiment and findings of this thesis are also presented. Finally, I propose further research and further design directions.

8 DISCUSSION

This chapter deals with interpreting and discussing the findings of the main experiment, as presented in the evaluation chapter. The discussion chapter is divided into four sections: The first section is a brief summarization of the main findings of the evaluation. The second section is a discussion of the findings. The third section is a discussion of the limitations of the findings and experiment of this thesis. The final section is a discussion of further research and design directions.

8.1 Summary of findings

Three hypotheses were put forth in this thesis. The hypotheses stated that playing HeartBee would elicit higher levels of flow, elicit higher levels of continuation desire, and lead to more time spent training than using HeartRate+ Coherence Pro (HRCP). All three hypotheses were confirmed.

Three secondary hypotheses were also put forth in the thesis. The hypotheses stated that the three measurements of an engaging player experience - flow, continuation desire, and time spent - would be correlated with each other. The secondary hypotheses were somewhat confirmed: The correlations were evaluated from four perspectives - while playing HeartBee, during the usage of HRCP, a combined measure of the two, and the change from HRCP to HeartBee. The findings confirmed all three hypotheses during gameplay of HeartBee and as a combined measure of the two. During the usage of HRCP, however, only continuation desire and flow proved to be correlated. The strongest correlation was found between Δcontinuation desire and Δflow where HRCP was considered a baseline measurement.

A focused analysis of the nine components of a flow experience while playing HeartBee revealed that *a challenging activity that requires skill*, *clear goals*, *clear feedback*, and *concentration on the task at hand* were the highest scored components whereas *merging of action and awareness* and *paradox of control* were the lowest scored components.

During usage of HRCP, a challenging activity that requires skills, concentration on the task at hand, and loss of self-consciousness scored higher than the remaining components.

An evaluation of the qualitative data gathered through the basic game ESQ after playing HeartBee revealed that positive affect, progression, and interfacing were the most commonly reported engagement triggers during gameplay whereas interfacing was also the most commonly reported disengagement trigger. This was mainly due to technical problems with the sensor used.

An evaluation of the qualitative data gathered after using HRCP revealed that progression, positive affect, and interfacing were the most commonly reported engagement triggers whereas negative affect, a lack of objectives, and a lack of progression were the most commonly reported disengagement triggers.

8.2 Discussion of findings

Flow, continuation desire, and time spent on a game were all concepts that were described as part of an engaging player experience. The main findings of this thesis indicate that HeartBee was a more engaging experience than HRCP.

8.3 Flow

Csikszentmihalyi (1990) describes a flow experience as an experience that encompasses the nine components of flow. To understand why people experienced higher levels of flow while playing HeartBee than while using HRCP, I investigate how the two applications satisfy the nine components and triangulate with the individual flow component scores as well as the reported qualitative feedback. In general, the flow scores for the two applications were quite similar.

For both HeartBee and HRCP, the participants scored a challenging activity that requires skills high. This suggests that both applications presented a challenge level that was just right for the participants. A challenge too difficult would lead to anxiety whereas a challenge too easy would lead to boredom (Csikszentmihalyi, Flow: The psychology of optimal experience, 1990). For HRCP, however, 12 participants (39%) reported that they got bored during the experience. This statement supports the claims by Wollmann et al. (2016) who state that breathing exercises are regarded as boring. In comparison, only one participant reported boredom during HeartBee. This suggests that HeartBee should score higher in the first component of flow than HRCP. The reason for why it did not can be explained by the component being limited to evaluation from one item on the SFSS compared to four items on the FSS (Jackson, Martin, & Eklund, 2008). Indeed, the question asks participant to rate the statement "*I felt I was competent enough to meet the demands of the situation*" which I argue is more related to the skills being high than the balance between skill and challenge being even. The simplification of the components into single items is one of the primary shortcomings of using SFSS compared to FSS.

For both applications, the merging of action and awareness was given an average rating. This can possibly be explained by the new way of interaction that the participants were unfamiliar with. The interaction requires a strong awareness on the breath and for many it can feel unnatural breathing at that rate, which can break this flow condition.

The flow score for clear goals was higher for HeartBee than it was for HRCP. This is also evident from the qualitative responses indicating that participants lacked objectives during HRCP. A lack of objectives is also identified as a disengagement trigger by Schoenau-Fog (2011b) and can therefore also lead to boredom. For HeartBee, participants had a strong sense of what they wanted to achieve, e.g. better their own score, beat their friends, or even win the game.

The flow scores for clear feedback were also rated higher for HeartBee than they were for HRCP. For both applications I argue that there was clear explicit feedback in the form of HR. Csikszentmihalyi argues that clear feedback should be in relation to evaluating the goal that one is trying to achieve (Csikszentmihalyi, Flow: The psychology of optimal experience, 1990). In relation to clear goals, a lack of clear feedback can then be attributed to a lack of goals within the HRCP application and especially a lack of feedback towards reaching that goal. Csikszentmihalyi states that one must be able to recognize and gauge the feedback and goals to enjoy the activity. In the sense of HeartBee, the feedback is made very explicit in the shape of a score system as well as how the HR affects the movement of the bee. The higher score for clear feedback can therefore be attributed to the transformation of an otherwise abstract topic, HR and HRV, into a more understandable topic through the game elements. This validates the findings of Wollmann et al. (2016) who suggest that transforming HR data into visualization elements and gamification leads to higher levels of engagement.

The flow scores for concentration on the task at hand were high for both applications, though slightly higher for HeartBee. Csikszentmihalyi defines concentration on the task at hand as a complete focus of attention on the task at hand (Csikszentmihalyi, Flow: The psychology of optimal experience, 1990). This definition is very similar to the definition of immersion by Ermi and Mäyrä (2005). According to their definition, immersion is a three-dimensional construct consisting of challenge-based immersion, sensory-based immersion, and imaginative immersion. Both applications lend themselves to challenge-based immersion through learning to control one's own HR through breathing. Sensory-based immersion, however, was only identified in HeartBee through the audio-visual landscape. Three participants identified this as an engagement trigger. In comparison, none identified the audio-visual landscape as a trigger for engagement in HRCP. Imaginative immersion was not identified in any of the applications and should not be applicable due to the lack of characters and story in both. As such, HeartBee should lend itself better for concentration on the task at hand as it fulfils two dimensions of immersion. Both applications, however, scored high in this flow component and seem to be subject of a ceiling effect, i.e. that differences cannot be detected due to them both scoring so high.

The flow scores for paradox of control were higher for HRCP than they were for HeartBee. I attribute this difference to the interfacing related issues that arose from using the Empatica E4 wristband for controlling HeartBee. Four participants specifically mentioned that the sensor disconnected during gameplay and observations of the test showed that this number was much higher. In comparison, no participant reported problems or were observed to have problems with registering heartbeats during HRCP. This can be due to the algorithm of HRCP being better at detecting heartbeats. It can also be a result of the participants feeling a lack of feedback, i.e. they will not report problems because they do not feel them, whereas in HeartBee, they will, as it directly affects their score. In any case, I argue that the flow scores for paradox of control while playing HeartBee would be higher with a better algorithm or device for registering heartbeats.

The flow scores for loss of self-consciousness were high for both applications, though slightly higher for HRCP. The loss of self-consciousness is highly related to the concentration on the task at hand component and is therefore also subject to ceiling effects. A reason for why it was scored slightly lower in HeartBee could be the interfacing problems reported for the paradox of control which could break their attention and make them become aware of the surroundings.

The flow scores for transformation of time were average for HRCP but higher for HeartBee. A reason for why it scored lower in HRCP could be that a timer was included in the UI of the application, constantly reminding them how long they spent. There is no timer included in the UI of HeartBee. After completing the experiment, some players reported that the transformation of time was very hard to evaluate after playing HeartBee as they had no idea how long they had been playing for. When told, most of them were surprised and thought they spent an equal amount of time on the two applications. The results of this component are therefore hard to interpret, and it is possible that participants should be informed about how long they spent before answering this question.

The flow scores for autotelic experience were average for HRCB but higher for HeartBee. The autotelic experience relates to how intrinsically rewarding an activity is (Csikszentmihalyi, Flow: The psychology of optimal experience, 1990). In the sense of HRV biofeedback applications, the intrinsic reward could be the enjoyment of breathing at the specified pace. For both applications, three participants responded that it was intrinsically rewarding, e.g. through being relaxing. In addition to that, some participants reported that they were interested in learning to control their HR through breathing. One thing that separates the two applications is the lack of audio-visuals in HRCB compared to HeartBee. This is possibly why HeartBee scored higher. Another reason could be the phrasing of the question, i.e. "*I found the experience extremely rewarding*" which does not isolate intrinsic rewards from extrinsic rewards. In addition to that, Csikszentmihalyi state that few activities are solely autotelic, but rather a combination of autotelic and exotelic (done for external reasons, e.g. to take part in an experiment) (Csikszentmihalyi, Flow: The psychology of optimal experience, 1990).

8.4 Continuation desire

The desire to continue playing HeartBee was reported to be higher than the desire to continue using HRCP. This can be contributed to how well the two applications lend themselves to the conceptual categories presented in the OA3 framework (Schoenau-Fog, 2011b). Indeed, 36 statements were identified as being related to triggers of engagements in HeartBee. In comparison, 16 statements were identified as being related to engagement in HRCB. The main reason for why participants had a higher desire to continue playing HeartBee can be attributed to the sense of progression, e.g. them wanting to beat their score. Furthermore, there were no reports of missing an objective. In addition to that, participants reported different accomplishments (completion and achievement) as triggers for engagement in HeartBee, something that was lacking from HRCP.

The findings support the findings of Wollmann et al. (2016) which state that HRV biofeedback games can make people engaged in HRV biofeedback training. Furthermore, their findings suggest that even higher levels of engagement can be reached by including more mechanics into the game. The trade-off, however, is a reduced effect on the training aspects of HRV biofeedback.

8.5 Time spent

The time participants spent playing HeartBee was much higher than the time participants spent using HRCB. The time spent was capped at 900 seconds (15 minutes). 16 participants (52%) played HeartBee until they met that time cap. In comparison, no participant used HRCB for 900 seconds. In fact, the longest time spent using HRCB was 513 seconds.

Wu et al. (2017) state that time spent is a strong indicator for engagement. Johnson et al. (2016) state that time spent is associated with engagement in video games. According to their statements, and according to the correlations found in this thesis, participants were much more engaged in HeartBee than in HRCB.

The time spent playing HeartBee during this thesis was also longer than the time spent during pre-thesis HeartBee, in which participants were forced to play for 600 seconds (10 minutes) (Brejl, 2016). The responses back then suggested that playing HeartBee for 600 seconds was sufficient and they reached boredom after that. The increased time spent during this version of HeartBee then suggests that the new design has improved the game from an engagement perspective. The improvements of the game can possibly be attributed to the upgrade in sensor (though still troublesome) and general improvements such as improved tutorial and upgraded UI. The upgraded high score system especially was able to provide objectives for the players, e.g. beat their friends, whereas the old high score system had no names connected to the scores.

From an HRV biofeedback training perspective, more training should lead to better results. The findings of Prinsloo et al. (2010) showed that 10 minutes of biofeedback training led to faster reaction time, less mistakes, and stronger consistency in answers during a Stroop task. The findings of Paul & Garg (2012) showed that 20 minutes of daily HRV biofeedback training for 10 consecutive days led to increased HRV, reduced anxiety traits, and improved basketball skills. In my previous project I proved that playing pre-thesis HeartBee for 10 minutes improved short term resting HRV (Brejl, 2016). 26 participants (84%) in the current thesis played HeartBee for a minimum of 10 minutes. In comparison, no participant used HRCB for 10 minutes. This strongly suggests that HeartBee is effective at engaging people in HRV biofeedback training.

8.6 Correlation between flow, continuation desire, and time spent

From analysing the results of a flow experience, it becomes evident that the conceptual categories of the OA3 framework can be applied to express components of a flow experience. Indeed, Schoenau-Fog states that player engagement is related to a range of experiences, including flow (Schoenau-Fog, 2011b). This statement was validated through correlating the reported flow scores with the reported continuation desire scores. Especially the change in flow from one condition to another was related to the change in continuation desire from one condition to another. Furthermore, time spent on the application was related to both continuation desire and flow, which validates the statements of Nacke & Drachen (2011) who states that time spent is closely related to psychological concepts such as flow. The relationship between time spent and player engagement, both from a continuation desire standpoint and a flow standpoint, also validate the findings of Johnson et al. (2016) who state that time spent is associated with competence, autonomy, and relatedness of self-determination theory. Wu et al.

(2017) state that time spent is a strong indicator of engagement when watching videos. The relationship established in this thesis extends that to video games.

Time spent, however, is only a measure and not a cause for engagement and does therefore not warrant increased engagement or flow.

With flow being expressed as the optimal experience and continuation desire being a measure of player engagement regardless of whether all flow conditions are fulfilled, I argue that the three measurements are related in a hierarchy in the following way:

A flow experience will always lead to a desire to continue as the experience is considered optimal. In addition, a flow experience leads to more time spent.

Continuation desire leads to more time spent but does not necessarily lead to a flow experience.

Time spent is a measure of a flow experience and continuation desire but is not able to produce the two by itself (e.g. forcing someone to do a task for an extended amount of time does not make them reach a flow state or want to continue doing it).

8.7 Limitations of findings

In this section I discuss the limitations of the experiment and findings of the thesis.

HeartBee is developed as a HRV biofeedback game. The purpose of the game is thus to increase the HRV of its players. This thesis does not measure HRV of its participants as it builds upon the findings of Brejl (2016). As such, the present thesis cannot be used as a basis for stating that HeartBee increases HRV in its players. I argue, however, that the changes in design applied to HeartBee does not change the underlying game in a way that would change the HRV outcome of its players compared to pre-thesis HeartBee. The underlying mechanic of the game, controlling the bee with the HR through slow deep breathing, has not been changed. The main reason that HRV was not measured during this thesis was that it would introduce a lot of variables that would have to accounted for (e.g. no smoking, eating, or physical exercise before the experiment) and the test conditions would not be reversible, i.e. a participant would not be able to both play HeartBee and use HRCP as they would increase their knowledge and HRV through both applications. Furthermore, measuring HRV should be done through resting periods of e.g. five minutes before and after the experiment which would make the experiment take much longer to complete.

In relation to HRV, another limitation of this thesis is the method of measurement as well as the device that was chosen for measuring it. While it has been proven to be relatively valid and reliable, PPG is still suboptimal compared to ECG for measuring HR. For playing the game, however, I argue that there is no need for exact measurements if players are rewarded for performing the correct exercises. The intrusiveness of ECG equipment could also make the game less appealing and break flow for the players.

The device chosen for measurement, the Empatica E4 wristband, posed problems in the sense of losing players' heartbeats. This made players unable to control the bee at times which violated the paradox of control component of flow and thus possibly broke their flow experience. The flow scores after playing HeartBee are therefore expected to be even higher than already reported if participants used a better device. The E4 was chosen because its measurements had been validated through empirical studies. The camera-based PPG used by HRCP seemed to register more heartbeats. However, it also accepted more "wrong" heartbeats from noise than the E4.

The subjective measurements of flow and continuation desire was done by using Likert scales with values from 1 to 5 and 1 to 7, respectively. The findings of this thesis suggest a strong ceiling effect, i.e. being unable to detect differences because both conditions receive a maximum score. A broader spectrum of selection, e.g. through visual analogue scales (VAS), might provide a greater insight into the differences between the two conditions.

The measurement of continuation desire for this thesis was based only on the post-game part of the basic game ESQ which measures a participant's desire to try the game again. It can be argued that participants desire to try again should be low in an experiment where they are free to stop the experiment at any time they want, as they might stop it because of losing engagement. When triangulated with both flow and time, however, the measurement of desire to want to try again likely measures engagement. The other two sections of the basic game ESQ were disregarded - the pre-test measurement because the motivation was affected by partaking in an experiment, and the during measurement because it would interfere with the experimental design.

The measurement of flow was limited to only one Likert item per flow component because of using SFSS. This was chosen to limit the amount of questions that participants would have to answer during the experiment. The findings, however, suggest that the limited amount of Likert items may give a biased view of the whole flow experience. Through triangulation with the other measurements, the overall results of this thesis, i.e. higher levels of flow during HeartBee than HRCP, seem to be consistent and valid.

Finally, for both applications, participants could stop the experience at any time they wanted. There is a risk that the participants are stopping the experience while they still enjoy it, or right as they do not, and thus rate their flow and desire to try again higher than if they had been forced to continue with the activity for a longer amount of time. The time spent, then, when presented with flow scores and desire to try again, gives a richer picture of player engagement during the experiment and suggests even stronger results than when considered isolated.

8.8 Further research

A single session of HRV biofeedback training is not likely to have huge beneficial results for an extended amount of time. Multiple sessions over a long period, on the other hand, might have, as reported by Paul & Garg (2012). In this thesis, I showed that a HRV biofeedback game was more engaging than a non-game HRV biofeedback application. The results were limited to participants only using the applications once. Further research should look at if a HRV biofeedback game can keep retention of the players over an extended time and if the player engagement will stay high. Optimally, it should be compared to a non-game HRV biofeedback application.

The game presented in this thesis was able to hold 16 participants (52%) engaged for the full 15 minutes of the experiment, after which they were forcefully stopped. It would be interesting to let players play for longer durations to find out when players generally stop playing the game.

Another future direction is to look at the physiological and psychological effects of playing a HRV biofeedback game over a longer period. My previous project (Brejl, 2016) assessed the physiological short-term effect. To my knowledge, no study has looked at the long-term effects of biofeedback games.

The demographics of the participants for this thesis were relatively narrow. The participants represented the general mHealth user as explained by Carroll et al. (2017). The findings of this thesis should therefore be tested against different target groups. When doing so, one could also take player types into consideration with the purpose of finding out who it appeals to.

Finally, the findings of this thesis suggest a relationship between flow, continuation desire, and time spent playing. These findings should be verified through other studies, e.g. player experience studies not necessarily related to biofeedback. This could expand the understanding of player experience and help define player engagement. Furthermore, time spent proved to be a strong indicator of engagement in this project and should be considered as a variable for future measurements due to its objective nature, its non-intrusiveness, and its ease of application.

8.9 Further design

HeartBee is, despite its multiple design iterations, still a prototype. In this section I describe how the design should be propelled forward to reach the market.

One of the limitations of HeartBee is that it relies on the Empatica E4 wristband, a device developed mainly for research. Furthermore, the algorithms used by the device for detecting heartbeats proved to be too conservative, i.e. the device often missed beats.

A next step in designing HeartBee would therefore be to find a sensor that registered more (correct) heartbeats and at the same time reach a broader target group. The Samsung Gear S3 was mentioned in the design chapter as a possible device for measuring the HR. It was disregarded due to a lack of empirical validation. Future design steps could therefore be to investigate the validity of the measurements of the device and make the game playable with it. The idea of controlling games on a phone from a smartwatch could also excite users and help propel the sales of the smartwatch.

Another idea is to use the camera-based PPG, as was used in HRCP. Camera-based PPG measures HR directly from the camera of the smartphone by having the user cover the phone with a finger. The camera can detect small changes in the colours of the finger which can be used to extract the HR, much like standard PPG. This method is highly desirable as it would allow users to play the game without needing third party hardware.

HeartBee should also be expanded to iOS. Camera-based PPG would make this transition seamless but extending the application to work with an Apple watch is another option.

In any case, a future design of HeartBee could lend itself to be played from a variety of sensors. Users with no third-party hardware could play it with camera-based PPG whereas users who own Bluetooth enabled HR sensors, both PPG and ECG, could utilize those, while smartwatch owners could use their smartwatch.

Another step for improving the interfacing aspect of HeartBee is to find players' individual resonance frequency instead of using a standard of 0.1Hz. Alternatively, the game could let players change the desired frequency freely, within limits.

Other than improving the interfacing of the game, future design iterations should expand the environment through adding different levels (e.g. an underwater level or an outer space level). This could introduce some variation in the game which could help engage players for even longer. Furthermore, the design should be guided by the OA3 framework and consider including more of the conceptual components to make it appealing to a larger target group.

9 CONCLUSION

The main hypothesis of this thesis was that a game based mHealth application, HeartBee, the output from a previous project, could increase user experience during heart variability (HRV) biofeedback training. To accommodate this, HeartBee was made mobile and the game design updated with the player experience in mind.

HeartBee was compared to a non-game HRV biofeedback application, HRCP, through user testing.

The findings confirm the main hypothesis: Participants of the study reported higher levels of flow ($p = 0.04, r = 0.38$) and continuation desire ($p < 0.001, r = 0.51$) after playing HeartBee than they did after using the non-game HRV biofeedback application, HRCP. Furthermore, participants on their own spent more time playing HeartBee than they did using HRCP ($p < 0.001, r = 0.68$).

For an mHealth application to have true value, it is a prerequisite that the usage of it is embraced by its users. The work presented in this thesis shows the potential of utilizing the sensors from wearable technologies and smartphones to develop engaging mHealth applications through serious games. By properly applying game concepts and elements, developers can engage their users while nudging them towards bettering their own health.

The tests conducted as part of this thesis only let test subjects play the game once, and the impact on HRV was not tested for. Earlier work proved a positive impact on test subjects' HRV from using a previous version of HeartBee once for 10 minutes. Further research should investigate the long-term health benefits of playing HRV biofeedback games, using the games consistently over a longer period. After all, the positive impact on the user's physical condition and wellbeing is the main objective of an mHealth application of this type.

Another equally important aspect to investigate is if and how biofeedback games can keep user retention over extended periods of time thereby ensuring consistent long-term usage.

The main challenge for bringing HeartBee to the market is the sensor technology that was used for this thesis. Further design steps should therefore investigate commercialized alternatives for measuring the players' HR through e.g. wearable technologies or smartphone cameras.

10 REFERENCES

Bensound. (2018). *Royalty Free Music by Bensound*. Retrieved August 30, 2018, from Bensound: https://www.bensound.com/

Bjørner, T. (2015). *Qualitative Methods for Consumer Research.* Hans Reitzels Forlag.

Boyle, E. A., Connolly, T. M., Hainey, T., & Boyle, J. M. (2012, May). Engagement in digital entertainment games: A systematic review. *Computers in Human Behavior, 28*(3), 771-780. doi:10.1016/j.chb.2011.11.020

Brejl, R. K. (2016). *HeartBee: Development and Evaluation of a Personalized HRV Biofeedback Game.* Copenhagen: Unpublished.

Brockmyer, J. H., Fox, C. M., Curtiss, K. A., McBroom, E., Burkhart, K. M., & Pidruzny, J. N. (2009, July). The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing. *Journal of Experimental Social Psychology, 45*(4), 624-634. doi:10.1016/j.jesp.2009.02.016

Brom, C., Buchtová, M., Šisler, V., Děchtěrenko, F., Palme, R., & Glenk, L. M. (2014). Flow, social interaction anxiety and salivary cortisol responses in serious games: A quasi-experimental study. *Computers & Education, 79*, 69-100. doi:10.1016/j.compedu.2014.07.001

Brown, E., & Cairns, P. (2004). A grounded investigation of game immersion. *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, (pp. 1297-1300). Vienna, Austria. doi:10.1145/985921.986048

Carroll, J. K., Moorhead, A., Bond, R., Leblanc, W. G., Petrella, R. J., & Fiscella, K. (2017, April 19). Who Uses Mobile Phone Health Apps and Does Use Matter? A Secondary

Data Analytics Approach. *Journal of Medical Internet Research, e125*. doi:10.2196/jmir.5604

Chambers, A. S., & Allen, J. J. (2007, Ferbruary). Cardiac vagal control, emotion, psychopathology, and health. *Biological Psychology, 74*(2), 113-115. doi:10.1016/j.biopsycho.2006.09.004

Chen, J. (2007). Flow in games (and everything else). *Communications of the ACM*, *50*, pp. 31-34. doi:10.1145/1232743.1232769

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety*. San Francisco: Jossy-Bass Publishers.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York, NY, USA: Harper & Row.

Csikszentmihalyi, M. (2014). *Flow and the foundations of positive psychology : the collected works of Mihaly Csikszentmihalyi*. Berlin: Springer.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining "gamification". *MindTrek '11 Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments* (pp. 9-15). Tampere, Finland: ACM. doi:10.1145/2181037.2181040

Empatica. (2018). *E4 streaming server*. Retrieved August 30, 2018, from E4 for Developers: http://developer.empatica.com/windows-streaming-server.html

Empatica. (2018). *E4 wristband*. Retrieved August 31, 2018, from Empatica: https://www.empatica.com/en-eu/research/e4/

Empatica. (2018). *Empalink Sample Project*. Retrieved August 30, 2018, from Github: https://github.com/empatica/empalink-sample-project-android

Engeser, S., & Rheinberg, F. (2008, September). Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion, 32*(3), 158-172. doi:10.1007/s11031-008-9102-4

Ermi, L., & Mäyrä, F. (2005). Fundamental Components of the Gameplay Experience: Analysing Immersion. *Digital Games Research Conference 2005, Changing Views: Worlds in Play*, (pp. 15-27). Vancouver, British Columbia, Canada.

F.Thayer, J., & D.Lane, R. (2007, February). The role of vagal function in the risk for cardiovascular disease and mortality. *Biological Psychology*, *74*(2), 224-242. doi:10.1016/j.biopsycho.2005.11.013

Field, A., & Hole, G. (2003). *How to Design and Report Experiments.* London: SAGE.

Fox, C. M., & Brockmyer, J. H. (2013, July 1). The Development of the Game Engagement Questionnaire: A Measure of Engagement in Video Game Playing: Response to Reviews. *Interacting with Computers, 25*(4), 290-293. doi:10.1093/iwc/iwt003

Frank, D. L., Khorshid, L., Kiffer, J. F., Moravec, C. S., & McKee, M. G. (2010, June). Biofeedback in medicine: who, when, why and how? *Mental Health in Family Medicine, 7*, 85-91.

Gelineck, S., & Wang, J. (n.d.). *ID - 1: Introduction to Interaction Design.* Retrieved August 25, 2018, from Moodle: https://www.moodle.aau.dk/pluginfile.php/247642/mod_resource/content/1/ID_1_slides.pdf

Ginsberg, J. P., Berry, M. E., & Powell, D. A. (2010, July/August). CARDIAC COHERENCE AND POSTTRAUMATIC STRESS DISORDER IN COMBAT VETERANS. *Alternative Therapies in Health and Medicine, 16*(4), 52-60.

Google. (2018). *Android Studio*. Retrieved August 30, 2018, from Android Studio: https://developer.android.com/studio/

Google. (2018). *Breathing exercise*. Retrieved August 30, 2018, from Google: https://www.google.dk/search?ei=N-GIW9SaGIXcwQLsuJ2oCg&q=breathing+exercise&oq=breath&gs_l=psy-ab.3.0.35i39k1l2j0i67k1l6j0i203k1j0i67k1.955.1636.0.2719.5.3.2.0.0.0.128.263.2j1.3.0....0...1.1.64.psy-ab..0.5.275...0i10k1.0.mOSILCtNbPI

Guida, C. T. (n.d.). *dreamlo*. Retrieved August 30, 2018, from dreamlo: http://dreamlo.com/

HeartMath Institute. (2018). *HeartMath Institute*. Retrieved August 20, 2018, from HeartMath Institute: https://www.heartmath.org/

IJsselsteijn, W., Hoogen, W. v., Klimmt, C., Kort, Y. d., Lindley, C., Mathiak, K., . . . Vorderer, P. (2008). Measuring the Experience of Digital Game Enjoyment. *Proceedings of Measuring Behavior 2008*, (pp. 88-89). Maastricht, The Netherlands.

IJsselsteijn, W., Kort, Y. d., & Poels, K. (2013). *The Game Experience Questionnaire.* Eindhoven: Technische Universiteit Eindhoven.

IJsselsteijn, W., Kort, Y. d., Poels, A. K., & Bellotti, F. (2007). Characterising and measuring user experiences in digital games. *Title of host publicationProceedings of the International Conference on Advances in Computer Entertainment Technology (ACE 2007)*, (pp. 1-4). Salzburg, Austria.

Jackson, S. A., & Marsh, H. W. (1996). Development and Validation of a Scale to Measure Optimal Experience: The Flow State Scale. *Journal of Sport & Exercise Psychology, 18,* 17-35.

Jackson, S. A., Eklund, B., & Martin, A. (2010). *The FLOW Manual: The Manual for the Flow Scales.* Queensland, Australia: Mind Garden, Inc.

Jackson, S., Martin, A., & Eklund, R. (2008, October). Long and short measures of flow: the construct validity of the FSS-2, DFS-2, and new brief counterparts. *Journal of Sport & Exercise Psychology, 30*(5), 561-587.

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008, September). Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies, 66*(9), 641-661. doi:10.1016/j.ijhcs.2008.04.004

Johnson, D., Gardner, J., & Sweetser, P. (2016, October). Motivations for videogame play: Predictors of time spent playing. *Computers in Human Behavior, 63*, 805-812. doi:10.1016/j.chb.2016.06.028

Koster, R. (2004). *A Theory of Fun for Game Design.* Paraglyph Press.

Laborde, S., Mosley, E., & Thayer, J. F. (2017, February 20). Heart Rate Variability and Cardiac Vagal Tone in Psychophysiological Research – Recommendations for Experiment Planning, Data Analysis, and Data Reporting. *Frontiers in Psychology, 8*. doi:10.3389/fpsyg.2017.00213

Lane, R. D., McRae, K., Reinman, E. M., Chen, K., Ahern, G. L., & Thayer, J. F. (2009, January 1). Neural correlates of heart rate variability during emotion. *NeuroImage, 44*(1), 213-222. doi:10.1016/j.neuroimage.2008.07.056

Lehrer, P. M., & Gevirtz, R. (2014, July 21). Heart rate variability biofeedback: how and why does it work? *Frontiers in Psychology*. doi:10.3389/fpsyg.2014.00756

Lehrer, P., Vaschillo, E., & Vaschillo, B. (2000, Sep 25). Resonant frequency biofeedback training to increase cardiac variability: rationale and manual for training. *Applied psychophysiology and Biofeedback, 3,* 177-191.

Leiker, A. M., Miller, M., Brewer, L., Nelson, M., Siow, M., & Lohse, K. (2016). The Relationship Between Engagement and Neurophysiological Measures of Attention in Motion-Controlled Video Games: A Randomized Controlled Trial. *JMIR Serious Games, 4*(1). doi:10.2196/games.5460

Lister, C., West, J. H., Cannon, B., Sax, T., & Brodegard, D. (2014). Just a Fad? Gamification in Health and Fitness Apps. *JMIR Serious Games, 2*(2). doi:10.2196/games.3413

Lombard, M., & Ditton, T. (1997, September 1). At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication*, *3*(2). doi:10.1111/j.1083-6101.1997.tb00072.x

Manzano, Ö. d., Theorell, T., Harmat, L., & Ullén, F. (2010). The Psychophysiology of Flow During Piano Playing. *Emotion, 10*(3), 301-311. doi:10.1037/a0018432

McCarthy, C., Pradhan, N., Redpath, C., & Adler, A. (2016). Validation of the Empatica E4 wristband. *2016 IEEE EMBS International Student Conference (ISC).* Ottawa, ON, Canada: IEEE. doi:10.1109/EMBSISC.2016.7508621

McCraty, R., & Childre, D. (2010, July/August). COHERENCE: BRIDGING PERSONAL, SOCIAL, AND GLOBAL HEALTH. *Alternative Therapies in Health and Medicine, 16*(4), 10-24.

McCraty, R., & Shaffer, F. (2015, January 1). Heart Rate Variability: New Perspectives on Physiological Mechanisms, Assessment of Self-regulatory Capacity, and Health risk. *Global Advances in Health and Medicine*, 46-61. doi:10.7453/gahmj.2014.073

McCraty, R., Atkinson, M., Tiller, W. A., Rein, G., & Watkins, A. D. (1995, November 15). The effects of emotions on short-term power spectrum analysis of heart rate variability. *The American Journal of Cardiology, 76*(14), 1089-1093. doi:10.1016/S0002-9149(99)80309-9

McMahan, A. (2003). Chapter 3: Immersion, Engagement, and Presence. In M. J. Wolf, & B. Perron, *The Video Game Theory Reader* (pp. 67-68). New York, NY, USA: Routledge, Taylor & Francis Group.

Meersman, R. E., & Stein, P. K. (2007, February). Vagal modulation and aging. *Biological Psychology, 74*(2), 165-173. doi:10.1016/j.biopsycho.2006.04.008

Microsoft. (2018). *Visual Studio IDE*. Retrieved August 30, 2018, from Visual Studio: https://visualstudio.microsoft.com/vs/

Murray, J. H. (1997). *Hamlet on the Holodeck: The Future of Narrative in Cyberspace.* New York, NY, USA: The Free Press.

Nacke, L. E. (2015). Games User Research and Physiological Game Evaluation. In L. E. Nacke, *Game User Experience Evaluation.* Springer International Publishing Switzerland.

Nacke, L., & Drachen, A. (2011). Towards a Framework of Player Experience Research. *EPEX'11 at FDG 2011.*

Newzoo. (2018). *Newzoo Global Games Market Report 2018 Light Version.* Retrieved August 20, 2018, from Trend Reports:

https://resources.newzoo.com/hubfs/Reports/Newzoo_2018_Global_Games_Market_Report_Light.pdf?submissionGuid=4fedaba8-974b-4b40-aa7c-ee8c24c48e03

Nielsen, J. (1993, November). Iterative User-Interface Design. *Computer, 26*(11), 32-41. doi:10.1109/2.241424

Norman, K. L. (2013, July 1). GEQ (Game Engagement/Experience Questionnaire): A Review of Two Papers. *Interacting with Computers, 25*(4), 278-283. doi:10.1093/iwc/iwt009

Paul, M., & Garg, K. (2012, June). The effect of heart rate variability biofeedback on performance psychology of basketball players. *Applied psychophysiology and biofeedback, 37*(2), 131-144. doi:10.1007/s10484-012-9185-2.

Pelegris, P., Banitsas, K., Orbach, T., & Marias, K. (2010). A novel method to detect Heart Beat Rate using a mobile phone. *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology* (pp. 5488-5491). Buenos Aires, Argentina: IEEE. doi:10.1109/IEMBS.2010.5626580

Pohl, M. (2017). *325,000 mobile health apps available in 2017 – Android now the leading mHealth platform*. Retrieved August 20, 2018, from Research 2 Guidance: https://research2guidance.com/325000-mobile-health-apps-available-in-2017/

Polar Electro. (2018). *Polar H10 Heart Rate Sensor*. Retrieved August 30, 2018, from Polar: https://www.polar.com/en/products/accessories/H10_heart_rate_sensor

Prinsloo, G. E., Rauch, H. G., Lambert, M. I., Muench, F., Noakes, T. D., & Derman, W. E. (2010, October 17). The effect of short duration heart rate variability (HRV) biofeedback on cognitive performance during laboratory induced cognitive stress. *Applied Cognitive Psychology, 25*(5). doi:10.1002/acp.1750

Research 2 Guidance. (2018, March). *mHealth Developer Economics: How mHealth app publishers are monetizing their apps.* Retrieved August 20, 2018, from Research 2 Guidance: https://research2guidance.com/wp-content/uploads/2018/03/R2G-mHealth-Developer-Economics-2017-How-To-Monetize-Mobile-Health-Apps.pdf

Russell, J. A. (1980). A Circumplex Model of Affect. *Journal of Personality and Social Psychology, 39*(6), 1161-1178.

Samsung. (2018). *Gear S3*. Retrieved August 31, 2018, from Samsung: https://www.samsung.com/global/galaxy/gear-s3/

Schoenau-Fog, H. (2011a). Hooked! – Evaluating Engagement as Continuation Desire in Interactive Narratives. *ICIDS 2011: Interactive Storytelling* (pp. 219-239). Springer, Berlin, Heidelberg. doi:10.1007/978-3-642-25289-1_24

Schoenau-Fog, H. (2011b). The Player Engagement Process - An Exploration of Continuation Desire in Digital Games. *Digra 11 - Proceedings of the 2011 DiGRA International Conference: Think Design Play*.

Schoenau-Fog, H., Birke, A., & Reng, L. (2012). Evaluation of continuation desire as an iterative game development method. *MindTrek '12 Proceeding of the 16th International Academic MindTrek Conference* (pp. 241-243). Tampere, Finland: ACM. doi:10.1145/2393132.2393185

Schoenau-Fog, H., Reng, L., & Kofoed, L. (2015). Fabrication of Games and Learning: - A Purposive Game Production. *Proceedings of The 9th European Conference on Games-Based Learning*, (pp. 480-488).

Shaffer, F., McCraty, R., & Zerr, C. L. (2014, September 30). A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. *Frontiers in Psychology*. doi:10.3389/fpsyg.2014.01040

Sherry, J. L. (2006, January 10). Flow and Media Enjoyment. *Communication Theory, 14*(4). doi:10.1111/j.1468-2885.2004.tb00318.x

Silicon Laboratories. (2018). *Bluegiga BLED112 Bluetooth® Low Energy Dongle*. Retrieved August 30, 2018, from Silicon Labs: https://www.silabs.com/products/wireless/bluetooth/bluetooth-low-energy-modules/bled112-bluetooth-smart-dongle

Somatic Vision. (2018). *Somatic Vision*. Retrieved August 21, 2018, from Somatic Vision: http://www.somaticvision.com/

Steffen, P. R., Austin, T., DeBarros, A., & Brown, T. (2017, August 25). The Impact of Resonance Frequency Breathing on Measures of Heart Rate Variability, Blood Pressure, and Mood. *Frontiers in Public Health, 5*. doi:10.3389/fpubh.2017.00222

Sweetser, P., & Wyeth, P. (2005, July). GameFlow: a model for evaluating player enjoyment in games. *Computers in Entertainment (CIE) - Theoretical and Practical Computer Applications in Entertainment, 3*(3), 3-3. doi:10.1145/1077246.1077253

Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. (1996, March 1). Heart Rate Variability. *Circulation, 93*(5). doi:10.1161/circ.93.5.1043

Thayer, J. F., Hansen, A. L., Saus-Rose, E., & Johnsen, B. H. (2009, April). Heart Rate Variability, Prefrontal Neural Function, and Cognitive Performance: The Neurovisceral Integration Perspective on Self-regulation, Adaptation, and Health. *Annals of Behavioral Medicine, 37*(2), 141-153. doi:10.1007/s12160-009-9101-z

TheBiofeedback. (n.d.). *TheBiofeedback.com*. Retrieved August 21, 2018, from TheBiofeedback: http://thebiofeedback.com/

Tian, Y., Bian, Y., Han, P., Wang, P., Gao, F., & Chen, Y. (2017, July 4). Physiological Signal Analysis for Evaluating Flow during Playing of Computer Games of Varying Difficulty. *Frontiers in Psychology*, *8*. doi:10.3389/fpsyg.2017.01121

TipSense LLC. (2018). *AppCrawlr*. Retrieved August 30, 2018, from AppCrawlr: http://appcrawlr.com/

Ulrich, M., Keller, J., Hoenig, K., Waller, C., & Gröna, G. (2014, February 1). Neural correlates of experimentally induced flow experiences. *NeuroImage*, *86*, 194-202. doi:10.1016/j.neuroimage.2013.08.019

Unity Technologies. (2018). *TextMesh Pro*. Retrieved August 30, 2018, from Unity Asset Store: https://assetstore.unity.com/packages/essentials/beta-projects/textmesh-pro-84126

Unity Technologies. (2018). *Unity*. Retrieved Augusg 30, 2018, from Unity3D: https://unity3d.com/

Wollmann, T., Abtahi, F., Eghdam, A., Seoane, F., Lindecrantz, K., Haag, M., & Koch, S. (2016, August 25). User-Centred Design and Usability Evaluation of a Heart Rate Variability Biofeedback Game. *IEEE Acess*, *4*, 5531-5539. doi:10.1109/ACCESS.2016.2601882

Wu, S., Rizoiu, M.-A., & Xie, L. (2017). Beyond Views: Measuring and Predicting Engagement in Online Videos. *Proceedings of ICWSM 2018.* Retrieved August 30, 2018, from https://arxiv.org/abs/1709.02541

11 APPENDICES

This thesis includes only a digital appendix.

The digital appendix includes 5 main folders: AV production, Bachelor thesis, Experiments, Game, and Implementation.

The AV production folder contains a showcase of the thesis as an MP4 file.

The Bachelor thesis folder contains a copy of my old project, “HeartBee: Development and Evaluation of a Personalized HRV Biofeedback Game”.

The Game folder contains a Windows standalone build of HeartBee and an APK file for installing on an Android device.

The Implementation folder contains the Android Studio project and the Unity project.

The Experiments folder contains two sub folders: First iteration and Main Experiment

First Iteration contains the questionnaire of the first iteration experiment as well as the responses.

Main Experiment contains the questionnaires used during the main experiment as well as a Data folder. The Data folder includes the raw responses, the SPSS files used for data analysis, and a folder for data gathered during gameplay (IBI, name. score).