Utrecht University

MSc Thesis in Computing Science

Algorithmic Data Analysis Track

ICA-5964962

Multimodal Recognition of Freezing Behavior in Virtual Reality

Author:

Renzo Schindeler

Supervisor:
Dr. ir. R.W. Poppe

Second Examiner:
Dr. A.A.A. Qahtan

July 29, 2021

Department of Information and Computing Sciences

Utrecht University

Abstract

Freezing is a component of fear behavior and is an evolutionary defense mechanism found in humans and animals that can be characterized by reduced body motion, decreased heart rate, and increased muscle contractions. In previous literature, fear stimuli were often presented using pictorial stimuli, and behavior was measured intrusively. These factors could negatively influence to what extent natural behavior is shown. Therefore, in this thesis, a novel and less intrusive multimodal approach for recognizing freezing behavior is proposed, which uses virtual reality (VR) to present the stimuli and uses balance and gaze data for the recognition of fear behavior. In VR, the participants were approached by armed and unarmed characters, which were the fear-inducing and neutral stimuli, respectively.

Results of trials featuring a single type of stimuli showed that participants' gaze fixated more on fear-inducing elements of the stimuli, while gaze was more exploratory when neutral stimuli were shown. Additionally, when showing both types of stimuli simultaneously, the gaze fixated more on the fear-inducing stimuli than the neutral stimuli, which is in line with previous literature about fear and freezing behavior. When measuring body sway, it was found that some participants showed freezing behavior, whereas other participants showed opposing behavior. During the fear-inducing trials, adjustments in behavior were found once the stimuli were nearing the participants, suggesting freezing required a delay to manifest in the participants, though evidence of this was not strong.

These results suggest that the proposed method could be used to detect fear and freezing behavior and could lay the foundation for further behavioral research in freezing behavior. Additionally, gaze fixations appeared to be a good measurement of fear behavior.

Keywords: data science, psychology, fear, freezing, virtual reality, gaze tracking, body motion

Contents

Acknowledgments

I would like to express my gratitude to my supervisor, Ronald Poppe, who has guided me throughout this project. I would also like to thank Stefan van der Stigchel for guiding me with processing the gaze data. Furthermore, I would like to thank Muriel Hagenaars for guiding me with processing the balance data and with designing the experiments. Additionally, I would like to thank Greg Tamargo and Jack Brookes for their contributions with developing the *WiiBuddy* and *Unity Experiment Framework* Unity packages, respectively, and for providing support when using their assets during the development of the experiment software, which considerably sped up the development process. Finally, I would like to thank my family and friends for their love and support.

1 Introduction

We are often unaware that our body reveals a lot about our emotional state, even when we try to hide our emotions. The facial region alone gives a lot away, especially the eyes and facial expressions. Body posture and limb movements also show correlations to how we feel emotionally, for instance, based on how jerky our arm movements are or how closed our stance is. These behavioral cues, which we are unaware of, provide objective data about our emotional state. These are interesting to analyze from a psychological point of view. This data is more reliable and is less prone to subjectivity, which makes it more useful to use for the recognition of emotions. Fear is an emotion we can feel when we anticipate danger and is a defensive reaction mechanism. We have all felt it before when we watch a scary horror film or ride down a high and fast roller coaster at a theme park. Like other emotions, fear can be recognized based on behavioral changes, such as gestures, posture, and body movements.

1.1 Freezing Behavior

Fear is often accompanied by a characteristic behavior called *freezing*. It can be recognized in both animals and humans by reduced body motion or sway, decreased heart rate, and increased muscle contractions [1, 2]. Freezing behavior occurs as a response mechanism to a potential threat that could be close, and when no escape route is available [3]. It serves as a preparation process for an organism to act on the threat or to defend itself, often called the *fight-or-flight* response [2, 3, 4]. Freezing behavior has been extensively studied in many animal studies [5, 6]. However, it has been examined to a lesser extent in humans, despite being very characteristic for recognizing fear in humans. Therefore, a focus is put on the recognition of freezing behavior in this thesis.

1.2 Multimodal Behavior Analysis

Behavioral research requires reliably recorded and realistic behavior from participants in experiments. To track the participant's behavior, they can wear devices, such as sensors or markers, which are worn on the body or in a suit or straps. Such devices might be intrusive and make the participant more conscious of their behavior, which threatens the validity of the actions performed [7, 8]. Unobtrusive alternatives exist, such as capturing full-body movement [7] or gaze [9] using cameras. The down-sides of these approaches are that their accuracy and robustness are often lower compared to the sensor/marker approach, but due to their unobtrusive nature, they might be more preferable as natural behavior is more favorable in behavior research. A trade-off often exists between the level of intrusiveness and the quality of the captured data. Unobtrusive measuring devices on their own are not always good at recognizing emotions, but using them in conjunction can greatly improve the recognition accuracy in the automated analysis of emotions. Freezing behavior is hard to recognize, so it is of great use to combine various methods to recognize its characteristic behaviors, which will be discussed more in detail in Section 1.4.1. For instance, a stabilometric force platform could be used to measure the amount of body sway, motion capture techniques could provide limb-specific or full-body kinematics data, gaze trackers could be used to extract features from the behavior of the eyes, such as blinking rate, gaze patterns, pupil size, and eye openness. Gaze trackers could additionally provide data about what type of stimuli induce freezing behavior by tracking the participant's gaze behavior. A microphone could also be used to capture vocal behavior, as humans appear to have a higher vocal pitch when they are in fear [10].

1.2.1 Mitigating the Intrusiveness of Motion Capture

The suggestion of using motion capture data for the recognition of freezing behavior in the previous section may appear to be counter-intuitive, as the participant should wear sensors or markers and is regarded as intrusive. However, when virtual reality is used to present the visual stimuli, this should mitigate the issue to some extent, as the participant is less aware of wearing these markers. It makes it impossible for the participant to see the markers, as the virtual reality device obstructs vision from the surroundings. Presenting stimuli using virtual reality will be further discussed in Section 1.3. Motion capture also does not limit participants in their movements, compared to other much more intrusive methods such as electroencephalogram (EEG) headsets, which limit head movements (which will be discussed more in detail in Section 2.2.3).

1.2.2 Benefits of Using Multiple Modalities

Emotions are often ambiguously expressed, as some share similar behavior. For instance, fear and surprise both share behavioral cues such as having the eyes wide open, raised eyebrows, or a dropped-open mouth [11]. When these features are looked at individually, we might not be able to classify freezing behavior accurately. However, when the recorded data from multiple modalities is combined, a more accurate recognition of freezing behavior could be the result. Therefore, in this thesis a focus is put on a multimodal approach of recognizing freezing behavior.

1.3 Fear-Inducing Stimuli using Virtual Reality

Many behavioral studies make use of pictorial stimuli in their experiments to analyze emotional reactions [2, 3]. These stimuli are often sets of images or videos with contents of the three categories *neutral*, *positive* and *negative*. These are presented to participants using computer screens to induce an emotional response from the participants. For instance, participants can be shown pictures of angry faces or horror films to induce fear behavior. Although such methods show significant signs of fear, very few consider approaches with different types of visual stimuli where the participant can move more freely, which could result in more natural behavior. In this thesis, this will be attempted by using a *virtual reality* (VR) device to present visual stimuli. VR devices are stereoscopic head-mounted displays (HMDs) that include head motion tracking, which allow for immersive experiences by giving the user the illusion they are standing in a simulated virtual world through visual stimuli. Optionally, the participant could wear surround sound headphones for a deeper immersion, which also allow for audio stimuli during experiments. VR devices allow the user to look around and, in more recent applications, to even walk around, as opposed to presenting stimuli on computer screens, where the user is often sitting or standing still and is required to look in a single direction at a static screen. As stated before, behavioral research requires realistic behavior, which might not be possible when you restrict the user to stay in a single location, looking at a computer screen. Nowadays, VR devices have a much higher graphical fidelity, which allows for deeper immersive virtual scenes that appear very close to reality. This means that VR is highly applicable for behavioral studies. As participants are more immersed with the virtual world, more natural and trustworthy behavior will be shown [12]. More specifically, VR can be used for inducing fear and freezing behavior and could open up a pathway for future behavior research, where behavior can be analyzed by presenting stimuli in more ways possible than by using a traditional computer screen. Stimuli can be presented in the periphery of the participant in addition to being presented at the center of our attention. VR could solve the issue of participants feeling self-aware by distracting them from being recorded, as they are immersed in a virtual world, which should further improve data reliability. For research, this is also interesting, as you can experiment in a more controlled environment with limited distractions from the surroundings outside of the VR device. Researchers have much more control and can trigger various audio-visual cues and record behavior at these moments, which reduces challenges and time spent on data acquisition, trimming, and cleaning. Therefore, VR proves to be of great use in behavioral studies and the recognition of freezing behavior. For these reasons, VR will be used during the experiment of this thesis.

1.4 Automatic Recognition of Emotions

Classical behavior analysis of emotions has its challenges. It requires experts to manually label or recognize emotions. This process is time-consuming, relies on subjectivity, and the reliability and quality of the gathered data are relatively lower compared to more recent methods. With the rise of machine learning (ML), the recognition of emotions drastically sped up. The automatic recognition of emotions is of great relevance for many real-life applications. Socially aware computing systems which interact with human users and pick up their behavioral cues are better able to respond and appear as more natural, efficacious, and reliable [8], making them suitable for fields such as human-computer interaction (HCI).

1.4.1 Recognition of Freezing Behavior

Fear is hard to recognize as it is expressed subtly and often unconsciously. Fear is also often mixed up with surprise [13]. From an aggregate of various studies can be concluded that overall, humans tend to be universally worse at recognizing fear than other basic emotions from facial expressions [14], especially at lower emotional intensities [15]. However, fear behavior also does have various bodily correlates. Therefore, it would be interesting to look further into recognizing fear automatically by analyzing these. The automated recognition of fear has its applications in surveillance systems that scan large crowds of people and could aid in police interrogation processes or at customs and border controls. Such systems can recognize unusual behavior automatically and anticipate dangerous or criminal situations/activities (for instance, a thief who has just stolen something and is afraid of being caught or illegal trespassers). Other examples of applications where the automated recognition of freezing behavior specifically can be convenient are the treatment of phobias and the objective evaluation of horror films, video games, or attractions that induce fear as a form of entertainment. Therefore, the recognition of freezing behavior has many use-cases and applications, which have yet to be discovered.

1.5 Novel Approach Using Virtual Reality

Motion capture, gaze tracking, and virtual reality have already been active research areas for behavioral studies on their own. To the best of the author's knowledge, no attempts have been made yet using these modalities in conjunction for the recognition of freezing behavior in humans, so this thesis is the first attempt at this. Principally, VR has been applied in a limited manner in multimodal behavioral studies, so at the same time, this thesis tries to encourage the usage of VR in more research areas and behavioral studies. As stated before, VR can be used to present audio-visual fear-inducing stimuli in the periphery of the participant, while multiple modalities record the induced freezing behavior.

1.6 Research Questions

The main goal of this thesis is to improve the present ability to recognize freezing behavior in humans unobtrusively using multiple modalities. Three research questions have been formulated to accomplish this goal.

1.6.1 Research Question 1

> *What is the difference in measured effect in body motion and gaze behavior when fear-inducing stimuli are presented compared to neutral stimuli?*

Research question 1 has been posed to analyze what types of behavior can be found when fear is induced in humans. It will be analyzed what types of fear behavior humans show and whether they show freezing behavior in VR based on body motion and gaze behavior. Additionally, it will be analyzed if humans generally follow the same behavioral patterns or if individual differences occur. Research question 1 will be answered by gathering participants' body motion and gaze data during a VR experiment. Trials will be executed by presenting either fear-inducing or neutral stimuli during the trials. The data gathered during the neutral trials will serve as baseline data that will be compared statistically to the data gathered during the fear-inducing trials, to determine whether differences in measured effect exist. If required, further visual analyses will be performed.

1.6.2 Research Question 2

> *Does gaze centralize more towards one of the types of stimuli when fear-inducing and neutral stimuli are presented simultaneously?*

Research question 2 has been posed to analyze what types of behavior can be found when neutral and fear-inducing stimuli are presented simultaneously. It will be analyzed whether any of the types of stimuli will draw more attention. Research question 2 will be answered by gathering gaze data from further trials which feature both types of stimuli simultaneously. The time spent looking at both types of stimuli will be compared statistically.

1.6.3 Research Question 3

> *Do the measured effects in body motion and gaze behavior manifest with a delay?*

Research question 3 has been posed to analyze whether induced fear behavior can immediately be found when stimuli are presented or if it requires some time before becoming visible. Research question 3 will be answered by analyzing the trials featuring fear-inducing stimuli more closely. The trial data will be split up into intervals, and these intervals will be compared to the general behavior found throughout the trials to look for temporal changes in behavioral patterns.

1.6.4 Methodology

The research questions will be answered based on empirical research, as the proposed method is novel and limited related research exists about analyzing fear behavior in VR. Participants will be recruited for an experiment in VR in which they will be exposed to fear-inducing and neutral stimuli. During the experiments, multiple modalities will be used to record behavioral data from the participants. The exact experiment methodology will be discussed later in Chapter 3 but will be based on relevant literature, which will be discussed first in Chapter 2.

1.7 Challenges

Challenges exist when unobtrusive and multimodal behavior studies are performed. The data could be of poor quality, data points could be missing, and every device has to be optimally configured. Behavior could also be hard to recognize and is often person-specific. The usage of VR opens many possibilities for behavioral studies, but not without the introduction of some potential issues that should be taken into account when designing experiments. These challenges will now be discussed in more detail.

1.7.1 Person-Specific Behavior

Responsive behavior is typically person-specific, which poses challenges in distinguishing emotions from one another [11]. Overall, fear behavior is hard to recognize and can also show up differently, based on the intensity of felt fear. As stated earlier, fear also shares similar characteristics with surprise. The main characteristics of fear (discussed in more detail later in Section 2.2) do not all have to be present for someone to be in fear. For instance, a person worrying about an exam taking place in a few weeks does not show all symptoms but might have subtle and occasional sensations of fear. This makes the recognition of fear and freezing behavior challenging.

1.7.2 Data Quality

As stated earlier, the recorded data from the different modalities might not be of high quality. Unobtrusive recording devices are often more prone to noise in the recorded data [7]. Generally, recording devices tend to have at least a minimal amount of noise, which could negatively influence the quality of the data, and therefore classification accuracy. Other devices, such as gaze trackers, suffer from drift, which results in inaccurate pupil locations. Gaze trackers also lose sight of the pupil when the participant blinks, which might result in missing or incorrect data points for pupil location and size. Other devices which are connected wirelessly to the data gathering system might lose the signal for short intervals, which results in missing data points in the data. Besides that, such devices might be subject to signal noise. Therefore, a data pre-processing step is required, which should remove the data inconsistencies or fill the gaps of the missing data.

1.7.3 Data Alignment

Each modality might record data at a different frequency, or the data from the modalities might not align properly [7]. This poses a challenge that data cannot be compared or combined straight away. Therefore, an additional pre-processing step is required to align and synchronize the data tracks. Data tracks of different duration should be cropped to be of the same duration. Data tracks of different frequencies should be aligned by remapping the data to different frequencies using inter- or extrapolation. In other experimental setups, recordings might take place at different times or using varying equipment. In such cases, data can also not be combined straight away and requires a pre-processing step where data is normalized or adjusted first.

1.7.4 Virtual Reality

VR poses some challenges, which should be taken into account when designing experiments. For instance, the participant might suffer from motion sickness or fatigue based on the visual stimuli or the duration of the experiment. The experiment setup should also consider the limitations of movement in VR, as there is either limited or no option to move naturally in VR. For instance, the participant can get tangled with the cable connecting the VR device to a computer. The participant's behavior might be different and less natural, because of these restrictions. The participant might also show adjusted viewing behavior due to resolution and *field of view* (FOV) limitations of the VR device. Additionally, the lenses of VR devices have increased resolution and sharpness in the center and decrease farther away from the center, which could also play a role in adjusted gaze behavior.

Facial Feature Obstructions If the participant is required to wear a VR device during an experiment, it becomes challenging to track features from the face, as such devices cover a large part of the face. Gaze behavior could be tracked by VR solutions that have built-in eye trackers, which could mitigate this issue. It is also challenging to measure facial temperature due to VR devices' placement against the face, often with a leather or foam pad for comfort. These might cause heat build-up in the facial regions, which might cause inaccurate temperature data. This means VR has its benefits in multimodal experiments, but at the cost of not being able to track certain facial features.

1.8 Outline

The remainder of this thesis is structured as follows: Chapter 2 discusses related works, such as existing methods for the recognition of freezing behavior and literature about virtual reality. For these works, the benefits, shortcomings, and challenges will be discussed and compared. Chapter 3 discusses the methodology of the experiment, such as required materials and participants, session procedure, and required steps to gather valuable behavioral data. Chapter 4 provides results of the analyses performed on the data and discusses these results. Chapter 5 will draw conclusions from the results and discussions and will answer the research questions posed in Section 1.6. Additionally, it will discuss the merits and limitations of the proposed methods and lays the foundations for future research.

2 Related Work

In this chapter, related literature will be discussed. Section 2.1 will briefly discuss how emotions are expressed and some of the most popular models of emotions. Section 2.2 goes deeper into how fear behavior is characterized and discusses the various features we can analyze to recognize fear. It also discusses freezing behavior, which is an important type of behavior shown when in fear. Section 2.3 gives an overview of some methods from related literature to automatically recognize fear and freezing behavior. Section 2.4 discusses what VR is, how it could be applied in behavioral studies, and what the challenges are of using it.

2.1 Emotions

Humans are able to communicate by expressing and recognizing social signals and social behaviors. These are manifested through non-verbal behavioral cues, such as facial expressions, body postures, gestures, and vocal outbursts [8]. Social signals typically last a short time (milliseconds to minutes) compared to social behaviors (seconds to hours or days). Humans use social signals to express emotions, which communicate our mental state and feelings often accompanied by physiological changes. Emotions can be recognized by social behavioral cues analyzed from (hand) gestures, body posture, facial expressions, gaze behavior, the focus of attention, vocal behavior, and in social situations, the distance between individuals [8]. These cues are more or less universally recognizable by many cultures, meaning all humans recognize and express the same emotions similarly [14]. We have the ability to recognize quite well what emotion others are expressing or how they are feeling without expressing it using words, which makes us socially intelligent. It will now be discussed how we can classify emotions.

2.1.1 Classification of Emotions

Many attempts have been done at classifying and recognizing emotions in the field of psychology. Different views among experts exist, but most consider that at least five core emotions exist. Many experts also use synonyms for certain emotions which are shared among different models (e.g. *scared/afraid* or *happiness/joy*). Some of the most commonly used models in literature will now be discussed.

Ekman's Basic Emotions A popular, categorical approach is Ekman's list of basic emotions. This model states that the basic emotions can be categorized into the following six categories: *anger*, *disgust*, *fear*, *happiness*, *sadness* and *surprise* [16, 11]. These are distinguishable from each other based on particular characteristics, where each emotion is a discrete category. Later on, other emotions were added to this list, though many works are still based on the list of six categories.

Dimensional Models of Emotions Emotions can be spatially visualized and classified in various ways. A popular and easy to understand approach is Russell's *Circumplex Model of Emotion*, in which emotions can be plotted in a circular two-dimensional space with as vertical axis the measure of arousal/activation (*low–high*) and as horizontal axis the valence (*negative–positive*, a measure of (dis)pleasure) [17]. However, behavioral expressions typically belong in multiple emotion categories, and emotions could transition quite immediately between each other, blurring the lines of such a categorical approach and thus making the classification of expressions quite difficult [18]. Therefore, other models exist that take a more continuous approach. A popular alternative is Plutchik's *Wheel of Emotions*, which suggests eight bipolar basic emotions [19, 20]. This model considers that emotions can be of different intensities and can be mixed to form different emotions. At the center, emotions are at their highest intensity, and this intensity is reduced as you move farther away from the center. Both models are visualized in Figure 1 and Figure 2, respectively.

Facial Action Coding System (FACS) A popular method that uses facial behavior features is Ekman's *Facial Action Coding System* (FACS). FACS is used to taxonomize human facial muscle movements [22, 13]. This system can be used to manually code nearly any anatomically possible facial expression, provides an objective and comprehensive language for describing facial expressions, and can be applied for the recognition of basic emotions [8]. Individual or groups of facial muscle movements are assigned a numeric label, which is called an *action unit* (AU). For each emotion, the AUs are enumerated that are activated when that category of emotions is likely expressed. FACS has been proposed for various use-cases, such as the analysis of depression [23]. Annotating with FACS has the downside of being time-costly due to manual classification by experts, and it relied mostly on subjectivity. Therefore, this annotation process could be imprecise [8]. To solve this, FACS also has an automated form, which detects faces in videos and then determines the expressed emotion automatically.

Though, it should be noted that the AUs are purely descriptive for specific facial configurations and include no inferential labels, meaning FACS cannot be directly used for relating facial muscle movements to felt emotions [24]. It rather tells what facial muscles are *likely* seen at certain emotions. For instance, someone could hide their negative emotion by smiling, and FACS would likely classify this as happiness based on the activated AUs.

2.2 Behavioral Correlates of Fear

Fear is by many models of emotions considered as one of the basic emotions and is frequently categorized among the negative or unpleasant emotions, such as anger, disgust, and sadness. For instance, fear is plotted within the quadrant of high arousal and negative valence (displeasure) in Russell's circumplex model (visualized in bold in Figure 1) and has its own category in Plutchik's model (visualized within the dark green petal in Figure 2) and Ekman's list of basic emotions (Section 2.1.1) [17, 11, 19]. Fear can take place immediately following surprise and often oscillates with anger [11]. When in fear, one can experience various sensations, such as feeling cold (often in the hands) or shortness of breath (hyperventilation) [11]. Fear has numerous behavioral correlates, where freezing behavior is one of the fundamental characteristics in recognizing fear. These behavioral correlates will now be discussed in more detail.

2.2.1 Freezing

Freezing is used as the main outcome measure for fear in many animal studies and is closely associated with fear [4]. Freezing is one of the main behavioral responses to a potential threat, which can be recognized in both animals and humans [2, 3]. For instance, Ekman already found in his earlier works that both humans and monkeys show freezing behavior when they are in fear [11]. Freezing has an evolutional cause, as it acts as a defense mechanism, which could occur when no escape route is available or when a threat is near [5]. It

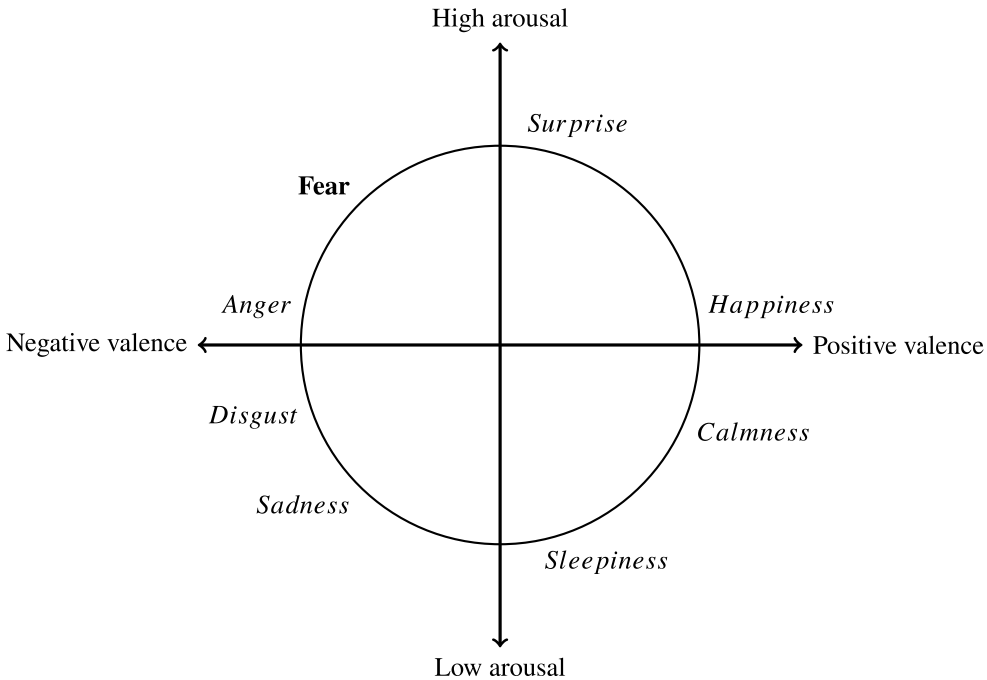


FIGURE 1: Russell's Circumflex Model of Emotion, with the measure of arousal as the vertical axis and the measure of valence as the horizontal axis. In this model, fear (in bold text) is plotted in the top left quadrant with relatively high arousal and negative valence.

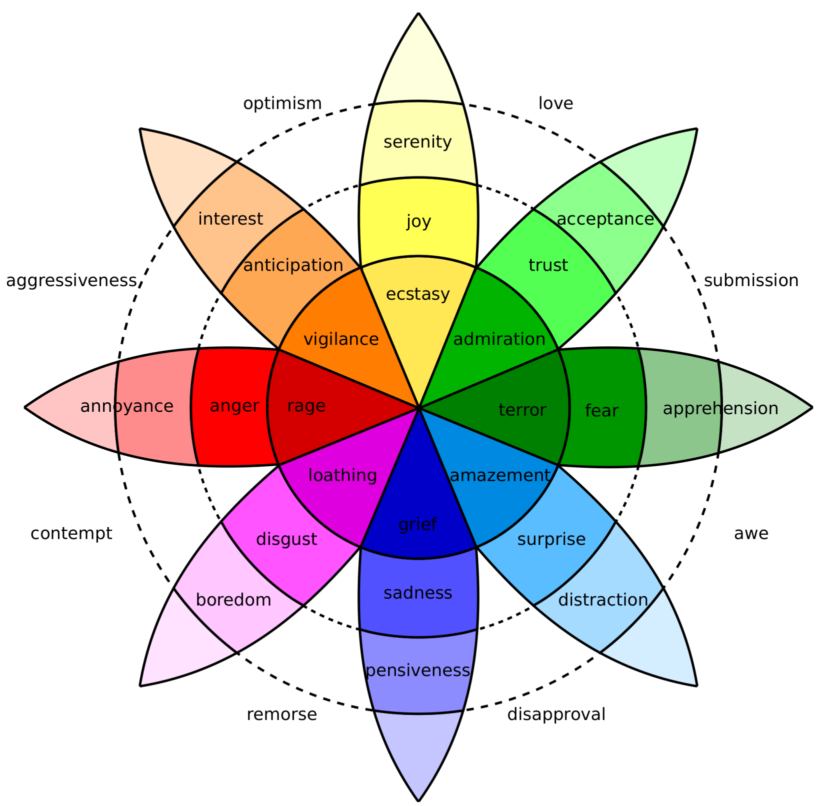


FIGURE 2: Plutchik's Wheel of Emotions, where emotions are bipolar, can be of different intensities and can be mixed. At the center, emotions are at their highest intensity, and this intensity is reduced as you move farther away from the center. Fear can be found in the dark green petal. Image source: Wikipedia [21]

serves as a survival mechanism that optimizes attentional processes and prepares the organism for action to defend itself [3, 4, 25, 26].

Behavioral Correlates of Freezing Freezing has various characteristic behavioral correlates. Various works found that the attention is directed towards the threat when freezing behavior occurs [27, 28, 29]. It is also characterized by reduced body motion, bradycardia (decreased heart rate), and increased muscle tonus (muscle contractions) [1, 2]. Muscle tonus is caused by blood flowing to the large muscles in the legs, which is supposed to prepare an organism to flee [11].

Freezing in Humans Freezing behavior has been broadly studied in animals in the past, but to a lesser extent in humans. Until the more recent years, models which describe the relations between animal and human freezing were also missing [4]. However, more research is being done into freezing in humans. In [2], research was done which determined that female humans appear to show freezing behavior, similar to animals. Significant reductions in heart rate and body motion were found when presented with negative stimuli in the form of angry faces as social threat cues, which induced fear as stated by [30], compared to happy or neutral faces. Research done by [3] confirms these findings and additionally found that participants who had endured aversive life events or traumatic experiences show increased reductions in heart rate. Human models on defense behavior are, to a large extent, based on animal models. However, a direct translation between animal and human models did not exist yet for freezing, until in [4] a translation model was proposed, which found consistencies between animal research and human research.

2.2.2 Facial Expressions of Fear

It is stated that the six basic emotions can be universally recognized and displayed from facial expressions [8]. Facial expressions are executed by contracting facial muscles, which cause movements of the facial skin and changes in the location and appearance of facial features [8]. The facial expressions of fear have the function to communicate a warning of potential harm with others or try to recruit others to help in dealing with a potential threat [11]. The most occurring facial expressions of fear are: raised and drawn together eyebrows, raised upper eyelids, tensed lower eyelids, lips stretched back towards ears, jaw dropped open, and a pulled back chin. Most of these facial expressions have AUs defined by FACS, which are listed in Table 1. More intense executions of these facial expressions showcase a greater intensity of fear. However, not all of these AUs have to be activated for someone to feel fear, and some of the AUs are shared with other emotions, which might make it challenging to recognize fear based on single AUs [11]. These challenges will be discussed in more detail in Section 2.3.2.

2.2.3 Physiological Correlates of Fear

Physiological behavior provides objective information about someone's emotional state, as this behavior cannot be hidden or consciously influenced. For instance, eye and gaze behavior can be used to obtain behavioral cues, as these play a crucial role in shaping social perceptions and showcasing emotions [8]. A downside of physiological features is that often they cannot be measured unobtrusively, as these features often require the measurement devices to be worn on the skin or body [8, 7].

AU Identifier	Facial Expression Description
1+2	Raised eyebrows
4	Eyebrows drawn/pulled together
5	Raised upper eyelids
7	Tensed lower eyelids
20	Lips stretched horizontally backwards towards ears
26	Jaw drop
×	Chin pulled backward

TABLE 1: The facial expressions of fear and their corresponding action units as described by FACS. The last item does not have an AU identifier but is often mentioned among other items in this list.

Pupil Dilation Previous literature has shown that pupil diameter increased significantly when participants were exposed to pleasant and unpleasant stimuli compared to neutral stimuli. This effect was visible for both pictorial [31] and audio [32] stimuli. In later works, a larger pupil size was related to an overall increase in arousal [33]. More recent works show stronger evidence that the pupil diameter is larger when exposed to negative content compared to positive content [33, 34, 35, 36]. In [27] a link between fear and pupil dilation is shown, where participants' pupils dilated when they were confronted with fear-inducing stimuli, such as angry faces, bodies, and scenes.

Blinking Behavior Various works show adjusted blinking behavior when participants anticipated danger or were shown fear-inducing stimuli. In [37] the eye blinking behavior of participants was analyzed when they anticipated electric shocks. The blinking amplitude was higher, and the latency was shorter in anticipation periods compared to safe periods. These findings were later confirmed by work done by [25], which additionally added that rapid eye closure is one of the most reliable components of the behavior of the startle reflex. Findings showed that blink magnitude for people afraid of spiders and snakes were higher when presented stimuli of their phobias. In research done by [9], the eye blinking rate (BR) was used in conjunction with other modalities to improve the classification performance of a classifier that was trained on data from participants who were shown horror films as fear-inducing stimuli. The change in BR did not appear to be significant enough and showed minimal correlations to other modalities, which could mean that BR is not a key element of detecting fear behavior in a multimodal setup.

Gaze Behavior Gaze behavior and *saccades*, small rapid jumps of fixations of the eye, are also key elements in recognizing fear behavior. Various previous works show how gaze behavior changes when fear-inducing stimuli are presented.

In research done by [25] it is found that people allocate more processing time to arousing, intense images regardless of valence. In [33] and [27] this is also found. When participants were confronted with a threat, participants' gaze was directed towards the threat. Threatening cues, especially angry signals induced arousal and were looked at longer than happy cues, according to [27]. In [33] attentional narrowing was observed following negative pictorial stimuli, as opposed to neutral and positive stimuli where this was not observed. Additionally, it was found that the attentional focus appears to narrow whenever individuals are encountering events of negative affective valence. In research done by [28] it is found that threatening stimuli are prioritized by the visual system and interfered with the executions of voluntary eye movements to other locations. Furthermore, it is found that when fear-conditioned and neutral stimuli were presented simultaneously, voluntary saccades were initiated faster toward the fear-conditioned stimuli compared to the neutral stimuli. More recent works also found these relations by using fear-conditioned stimuli. Fear-conditioned stimuli are relatively neutral stimuli combined with aversive events, which make the organism learn to predict these aversive events (also called Pavlovian fear conditioning) [38]. In research done by [39] and [29], participants were shown fear-conditioned stimuli, which made the participants anticipate an electrical shock. In the experiments of [39], trials were executed where participants were shown visual stimuli using a colored cross on a computer screen, which indicated whether they would get a painful electrical shock certainly, possibly, or certainly not. Results showed significantly fewer and longer fixations, which were closer to the center of the screen (where the fear stimulus, the cross, was shown) in trials where the shock would possibly occur. These findings show reduced visual exploration in participants who showed flight behavior. In [29], these freezing-like patterns of eye movements and centralization of gaze were also found when participants could escape from aversive stimulation, also in the form of electrical shocks.

From these findings, it can be summarized that when humans are exposed to fear-inducing stimuli, be it fear-conditioned or a direct threat, their natural gaze behavior is interrupted and adjusted by narrowing and fixating towards the fear stimuli.

Electroencephalogram Signals Electroencephalogram (EEG) signals are electrical brain activities, which can be measured using EEG devices or headsets. EEG devices are worn over the head by making electrodes touch the scalp's surface. Research shows that the ratio of the slow to fast wave powers of the EEG signal can be used to detect fear [9, 40]. EEG devices are widely accessible and have relatively low costs but are very sensitive to noise when the user moves the head and could be considered as intrusive.

Facial Temperature In research done by [9], it is observed that facial temperature data was among the most reliable features in evaluating fear unobtrusively. Experiments from these findings suggested a significant decrease in facial temperature when participants were shown a horror film after watching neutral content. Especially the right cheek appeared to have the most decline in temperature statistically.

Heart Rate The heart rate can be studied by analyzing an electrocardiogram (ECG), a graph of the electrical activity of the heart. A reduced heart rate is one of the key characteristics of freezing, as mentioned earlier in Section 2.2.1. In [25] a linear relation between heart rate and valence was found, which was found in experiments where unpleasant pictures were shown to participants. These induced decelerated heart rate during the viewing of negative stimuli. More recent work done by [39] also found a lower heart rate when fear-inducing stimuli were presented to humans.

2.2.4 Body Motions, Postures & Gestures of Fear

Research has shown that as much as 90% of body gestures are associated with speech and, in some cases, gestures are performed unconsciously, which are essential to track, as they provide honest information about a person's emotional state [8, 41]. An individual's emotional state can be recognized from a distance by analyzing posture or body movements, without the requirement of facial features [42]. This means the analysis of fear and freezing behavior can be done less obtrusively without requiring the physiological features discussed in the previous section.

Motoric Characteristics The motoric characteristics and positions of fear have already been described over a century ago by Darwin. He showed that when humans are in fear, the body contracts in an attempt to appear as small as possible and tries to protect itself [43, 44]. The head sinks between the shoulders by slouching or hunching, the individual becomes motionless or crouches down, the arms are thrown wildly over the head, and the arms violently protrude as if to push away. In more recent literature, a set of motor characteristics was used to predict basic emotions. Fear was considered an avoidance emotion and was predicted by enclosing and condensing the body to appear smaller or non-visible, to avoid battle, to avoid being hurt, and to defend vital organs, as well as by moving backward in space and retreating in the shape of the body (e.g. leaning backward with the torso). These extend the findings of Darwin [26]. These findings were in line with well-known responses to danger in the animal kingdom [4, 26].

Limb Movements Research performed by [45] experiments with recognizing emotions from arm and wrist movements by capturing joint movements for knocking and drinking behavior. Results show that simple arm movements are by themselves effective in communicating affect. In their experiments, the wrist kinematics (velocity, acceleration, and jerk) were tracked. It was found that the activation axis (arousal) was highly correlated to physical characteristics of the movement, such that greater activation (higher arousal) was related to greater magnitudes of velocity, acceleration, and jerk of the movement. In the case of fear, this means that wrist and arm kinematics are relatively lower compared to other emotions. Additionally, it is suggested that this relation could be found in other limb or joint movements.

Posturography Research has related a reduction of body sway/motion to freezing behavior [3, 2], as was mentioned earlier in Section 2.2.1. This was measured by showing participants images of facial expressions while standing on a stabilometric force platform. Experiments have shown that when participants are shown angry faces, they have reduced body sway and heart rate, which correlated to increased subjective anxiety scores. Such methods allow for unobtrusive analysis and appear to be significant enough for the classification of freezing behavior.

2.2.5 Vocal Behavior of Fear

Vocal behavior also changes based on someone's current emotional state. The main energy source in speech is the vibration of the vocal cords, from which the vibration rate determines the voice pitch (fundamental frequency or F0) [46]. Research shows that when in fear, our speech rate is much faster, the average pitch is much higher, voice quality is irregular, and articulation is more precise [10]. In social situations, energy bursts in the voice are present, such as shouting [8].

2.3 Automated Recognition

Much research has been done on the automatic recognition and classification of emotions. Classical approaches like FACS made use of manual labeling, while presently, we use machine learning techniques that automate these processes.

2.3.1 Measurement

Many of the fear and freezing correlates discussed in Section 2.2 can be measured automatically. This section will provide a non-exhaustive overview of some techniques from the literature and some commonly used devices which can record behavioral data. The findings are summarized in Table 3.

Full-Body Motion Various body motion tracking devices exist, which capture body kinematics data. These devices can be categorized by how they record data (using markers, inertial or vision data) and whether they offer full-body or single-movement capture data [7]. Marker-based approaches use a set of cameras to detect markers worn on the body. Inertial-based methods do not require cameras but measure movements on the body through sensors. Vision-based techniques make a digital volumetric estimation of the scene and the participants based on recordings from cameras only, thus being the most unobtrusive but also the most inaccurate.

A vision-based approach is the *Computerized Emotion Perception based on Posture* (CEPP) algorithm, which estimates someone's emotional label [47]. This algorithm makes use of an image capturing device to extract a silhouette of a user. From this silhouette, an edge image is generated, and a center point is computed. From these, the Euclidean distance is computed compared to the ground truth image. The ground truth's image with the lowest distance is considered the final class label. This technique is computationally fast and does not require the algorithm to be trained, but only looks at the overall body pose instead of segmenting the body and looking at individual body components' poses. A downside of this approach is that participants were requested to mimic a certain (exaggerated) emotion. This makes the technique unsuited for more realistic applications or more subtly expressed emotions. An alternative vision-based technique discussed in [48] mitigates this problem by using actors. They were instructed to act out an emotion without any guidance or instructions, resulting in the data gathered during this research being quasi-natural. The downside of acted emotions is still in place, but with increased data quality, diversity and cleanness. Their approach analyzes the influence of also including the lower body, as most research focuses on the upper body, which appears to show significant influence. Tests were done using three neural network (NN) architectures (CNNs, RNNs, and RNN-LSTM). Various datasets were generated (upper body with and without lower body, usage of position, and orientation data of joints). The RNN-LSTM appeared to perform best at recognizing emotions using only positional data of the skeleton. This technique worked well for simple gesture recognition. However, it had trouble recognizing complicated and non-repeatable expressions.

A technique discussed in [49] makes use of a single recording device, a *Microsoft Kinect*, to analyze the affective states of players of serious games. Kinematic, spatial, smoothness, symmetry, leaning, and distance-related features were recorded and fed to a deep learning classifier, which obtained high classification accuracy. Distance-related features appeared to have the largest recognition rate among the different groups of features.

In [7] a method called *Automated measurement and analysis of body motion* (AMAB) is discussed, which is a standardized methodology to analyze full-body motion quantitatively based on output from full-body motion tracking systems. This method can be used to analyze and compare data gathered from different skeletal motion capture setups.

Body Sway For the recognition of freezing, stabilometric platforms can be used to capture reduced body sway [3, 2]. Participants can stand on a platform without being required to wear any sensors/markers or for the setup to require any cameras, which makes them relatively unobtrusive. These devices can determine the center of pressure (COP) for participants, which are used to determine the body sway by looking at the frequency and magnitude of movements of the COP. These devices provide more informative and subtle information compared to using a scale [50]. However, they are often expensive and difficult to set up and transport. Nintendo's *Wii Balance Board* is a much cheaper, portable, and accessible solution, which is satisfactory for assessing standing balance compared to the clinical standards. This device has been widely used in various literature, for instance as an interaction device for VR [51]. However, it is designed as a profitable, low-cost video gaming controller. Thus, it has some limitations, such as the inability to assess force in the horizontal axes and the inability to be applied in situations that require high and rapid force. This means that, as long as the participant is only required to stand, this device is suitable for recognizing reduced body sway.

Facial Expressions Facial behavior plays a major role in shaping social perceptions and can be measured automatically, which can be done with the aid of visible-light and thermal cameras [8]. Beforehand, regions of interest (ROIs) should be defined, which tell what sections of the face should be measured or captured [9]. Machine learning techniques such as *adaptive boosting* (AdaBoost) [52] can be used to detect the facial regions and individual facial elements, such as eyes and nose, to track the ROIs. Alternative techniques consider facial-point- or facial-point-contour-tracking or an ensemble of classifiers [8].

Facial feature changes are detected by analyzing optical flow or by tracking facial feature points [8]. The optical flow approach has the advantage it does not require a facial feature extraction stage (thus, defining ROIs is not required). Additionally, it can be used to directly represent facial expressions, as it is expressed in velocity and can use flow information from the whole face. Optical flow techniques were popular in the final years of the 20th century [53, 54]. However, these techniques do have some limitations, such as the sensitivity to noise, occlusion, clutter, and changes in illumination [8]. Instead, various novel automatic facial expression recognition methods were found that tackle these issues by tracking facial feature points in image sequences, such as the mouth corners, lips, and eyebrows. In [55], automatic feature point tracking demonstrated high concurrent validity with human coding using FACS.

Facial Temperature In [9] facial temperature was one of the measured facial features in recognizing fear and was one of the most reliable fear correlates. This can be measured using a thermal camera. However, it is challenging to detect the regions of facial features using a thermal camera alone, because the textures of facial features are not distinct, and therefore not properly visible on the thermal image. A solution to this is to use a visible-light and thermal camera in conjunction. The visible-light camera detects the ROIs, which the thermal camera uses to acquire the facial temperatures at the required locations. As these cameras could have different parameters (intrinsic (focal length, lens distortion) and extrinsic (camera pose)), a geometric transformation matrix is used to map between pixel positions of the different image captures, which is also called camera calibration. However, this requires the participants to have their faces in the FOV of the cameras at all times, thus limiting freedom of head movements.

Eye & Gaze Features Eye features that are often used in the literature to classify emotions are pupil diameter, pupil position, blinking rate, fixation duration, saccade fixations, motion speed of the eye, and eye openness [56, 57, 9, 58, 12, 59, 60]. These features can be analyzed using gaze trackers, which provide reasonably accurate results with minimal gaze angle error when calibrated properly [12, 59, 60].

Before these features can be tracked, the eye and pupil must be located automatically first, which is done using various iris detection algorithms [61, 9, 62]. Image processing techniques exist, which extract the pupil diameter from the eye ROIs by detecting the iris inner and outer boundaries and removing noise, such as eyelashes and eyelids [62]. The blinking rate is obtained by considering a blink as a short time interval where the eye tracker cannot detect the pupils [9]. When fixations are defined as periods with stable gaze between saccades, gaze features such as saccades, fixations, and eye motion speed can be obtained by tracking the pupil's coordinates [39].

Gaze trackers can also be implemented by using webcams or front-face cameras of mobile devices, but these devices require the eyes to be within FOV and not be occluded. A solution is to make the participant wear a gaze tracker, but this might be intrusive and impact the behavior shown [63]. Therefore, VR solutions exist with built-in gaze trackers, such as the *HTC Vive Pro Eye*, which has a built-in *Tobii* eye tracker [64]. By using this device, gaze can be tracked to determine what objects or stimuli the participant is looking at in VR and for how long. In [12] a method is discussed to capture gaze in the *virtual environment* (VE) by using a VR device that has a built-in eye tracker. How this works, will be discussed later in Section 2.4.4.

A drawback of tracking eye behavior (such as pupil diameter) is that the pupil is highly sensitive to light. Additionally, there will be signal noise provoked by blinks and saccades, which means this data cannot be used directly for emotion recognition and requires a pre-processing step to remove these inconsistencies [35, 36]. Additionally, these devices need to be calibrated for each user, may suffer from drift, and in the case of VR devices with an eye-tracker built-in, the user is not allowed to move the VR device. Otherwise, the tracking might be offset from the calibration [12].

2.3.2 Interpretation of Fear

The automatic recognition of fear and freezing behavior is challenging. Fear and surprise have similar behavior correlates. It is also hard to distinguish between fear and anxiety, as both are threat-related emotions but are induced differently. Therefore, one should be cautious when trying to classify fear of freezing behavior automatically. These challenges will now be discussed in more detail.

Distinguishing Between Fear & Surprise A challenge is that fear is more difficult to recognize than other emotions, as it has less pronounced characteristics or characteristics that are shared with other emotions [11]. Surprise and fear often occur sequentially and have similar characteristics, which makes them hard to distinguish from one another. Usually, surprise has a short duration, while fear can span a longer time. The eyelids and eyebrows are of great importance to tell the two apart. The lower eyelids are more relaxed, and the upper eyelids are raised to a lesser extent in a surprised state. Additionally, the eyebrows are less drawn together in a surprised state. Though, raised eyebrows can appear ambiguous as this is often used as an emphasis sign when talking. Signs of surprise and fear can also be told apart by looking at someone's mouth and lips. Frequently, when someone's mouth is open, it is a sign of surprise, while in fear, the lips

Emotion Characteristic	**Fear**	**Anxiety**
Behavior direction	Threat avoidance	Approach threat
Brain activity region	Right hemispheric activity	Left frontal hemispheric activity
Cardiac activity	Reduced heart rate	Increased heart rate & palpitations
Threat specificity	Certain	Ambiguous & uncertain
Temporal focus	Present-focused	Future-focused
Duration of emotional experience	Short	Long
Body motion	Reduced body motion	Increased motor tension

TABLE 2: A non-exhaustive overview of the differences between the characteristics of fear and anxiety gathered from related literature.

are stretched back towards the ears. For both emotions, eyebrow movements can be absent, or signals can be mixed, so one should be cautious when focusing on single facial features.

Distinguishing Between Fear & Anxiety Fear and anxiety are both threat-related emotions but are induced differently [65, 4]. Fear occurs in response to actual threats, such as aversive stimuli or a predator. on the other hand, anxiety can be evoked by a potential or uncertain threat or based on a certain expectation. For instance, this could be a predator that has been encountered in the past. Additionally, a distinction is made between freezing or anxiety behavior situations, which is motivated by whether the threat is avoidable. Fear-motivated behavior is directed away from the threat, whereas anxiety-motivated behavior is directed towards the threat [66].

In a review done by [67] the relationship between anxiety and fear is analyzed. Results show that both do show only a moderate relationship, and they suggest that both are distinct emotions, despite some similarities. Both differ in behavior direction, neurobiological activations of the brain, heart rate, type of threat, temporal focus, and duration of felt emotion. Furthermore, anxiety could result in increased motor tension (e.g., foot-tapping), whereas fear results in reduced body motion [68]. An overview is given of the differences between both emotions gathered from literature in Table 2.

2.4 Virtual Reality

VR is a growing area of research, as its popularity has increased immensely over the last few years in areas such as architecture, medicine, and military and space simulators [69]. The core ideas of VR are immersion and interactivity with a virtual world [70]. VR devices are popular due to their constantly increasing affordability and accessibility. For instance, in [71] it is stated that the *HTC Vive* [64] proves to have sufficient tracking accuracy and precision and has relatively low latency and noise levels. These make it applicable in research done with far more expensive systems, despite having imperfections with estimating the true ground plane, which cause incorrect roll, pitch, and height measurements.

2.4.1 Optical Positional Tracking

Nowadays, various solutions exist that allow the user to move around more freely in a VE when wearing a VR device. Classical devices only allowed for looking around in the VE in a static position and could not track positional data. The newer devices use optical positional tracking, a method that tracks the device's location

Behavioral Feature	Example Measuring Device	Intrusive	Advantages	Disadvantages
Body motions, postures and gestures	Stabilometric platform	No	• Accurate and key indicator in detecting reduced body motion • Cheap solutions exist (e.g. Wii Balance Board)	• Can only track COP, so cannot track individual body part kinematics data
	Sensors and markers	Yes	• Can track individual body part kinematics data • Relatively high accuracy • Trackable from far away	• No occlusion allowed for marker-based methods • Requires multiple cameras for higher precision
	Cameras	No	• Estimation solutions exist which work with a single camera (e.g. webcam, Microsoft Kinect)	• Generally less accurate than sensors/markers approach • No occlusion allowed
Physiological behavior	Gaze tracker	Yes	• Relatively accurate gaze tracking/minimal gaze angle error • Can track rich features of fear	• Requires calibration per user • Suffers from drift • No head movements allowed
	EEG device	Yes	• Accurate in recognizing fear from brain activity	• Suffer from noise when the user moves its head
	Thermal camera	No	• Accurate in detecting reduced temperature in the facial region	• Difficult to track ROIs
	ECG/EKG sensor	Yes	• Accurate in detecting reduced heart rate	• Not applicable from far away or inaccurate portable solutions
Vocal behavior	Microphone	No	• Built-in in most VR solutions	• Vocal behavior might not always occur
Facial expressions	Cameras	No	• Labeling solutions exist (e.g. FACS)	• Facial expressions not always in line with felt emotion (not an objective measurement) • Face must be in FOV of the camera

TABLE 3: A non-exhaustive overview of the behavioral correlates of fear and freezing and their respective measuring devices. Additionally, advantages and disadvantages of the devices are listed.

in an environment based on image sensors and markers. Two types of optical positional tracking techniques which are often used at the time of writing are *outside-in* and *inside-out* tracking [72].

Outside-In Tracking Outside-in tracking uses other devices, such as cameras in stationary locations to track the position of markers on the VR device. This method is similar to how marker-based motion capture applications work [73]. These methods have high tracking accuracy and can offload the tracking work to the tracking devices. However, these methods require that the user is within the *field of view* (FOV) of the tracking devices. Another downside is that tracking accuracy decreases when the VR device is occluded.

Inside-Out Tracking Inside-out tracking uses cameras or sensors placed on the VR device itself, which determine its location in the environment. This method does not require external devices to track the location of the device, allowing for no restrictions in movement. However, these methods suffer in tracking accuracy and latency and have increased workload, as the computational work is done by the headset itself.

2.4.2 Virtual Reality Platforms

Consumer-grade VR devices exist for video game consoles, such as Sony's PlayStation VR [74]. Google Cardboard is Google's approach at bringing VR applications to consumers in an affordable way by turning any smartphone into a wearable VR device using a cardboard viewer, which you can use to watch 360-degree videos or play games [75, 76]. Thanks to the development of Valve's SteamVR platform [77], a large influx originated in the consumer-grade VR devices market for PC in recent years. This platform makes it possible to experience VR content on a great range of devices. Additionally, this platform enables the modular use of devices, which makes it relatively easy to combine or replace hardware from different manufactures. For developers, it has an easy-to-use *software development kit* (SDK) which makes developing for VR much more accessible. Popular game engines, such as Unity [78] and Unreal Engine [79] have native support for the development of cross-platform VR games using this platform.

2.4.3 Virtual Reality in Behavioral Research

Nowadays, VR devices allow for more immersive virtual scenes which are very close to reality. This means they can be used for behavioral studies. As participants are more immersed with the virtual world, more natural and trustworthy behavior will be shown [12]. VR can tackle the participant being self-aware during experiments by making the user feel immersed in a virtual world, which improves data reliability. For research, this is also interesting, as experiments can be done in a controlled environment. Researches have much more control and can trigger various audio-visual cues and record behavior at these moments, which reduces challenges and time spent on data acquisition, trimming, and cleaning.

Virtual Reality as a Treatment for Fear & Medical Conditions VR has been applied to treat various phobias and medical conditions. In [80] VR was applied to patients with peripheral vestibular dysfunction. Patients suffered from symptoms, such as *visual vertigo* (VV). A group of patients was exposed to a wide FOV dynamic VE while combined with vestibular exercises, while a different group took the same treatment without the use of VR. Significant improvements were shown for symptoms of VV when VR was included in the rehabilitation program. Subjective symptom scores of VV were greater when VR was combined with exercises.

More relevant to fear or freezing, in [81] VR is used as an exposure component for the treatment of panic disorders with *agoraphobia* (PDA), a phobia of entering open or crowded places or being outside of the home. The treatment of PDA requires the patient to undergo graded exposure to what they fear. Advantages of VR exposure are that the patient is in a safe and controlled environment, the patient is treated confidentially (rather than in a public location), the possibility of doing VR interoceptive exposure, and the cost-benefits of VR devices and treatment time.

In [82] participants suffering from *arachnophobia*, the fear of spiders, were treated with exposure sessions to spiders using VR. Various questionnaires taken pre- and post-test showed significant decreases in scores in fear, meaning VR could increase treatment effectiveness.

In [83] VR has been validated in experiments that study conditioned suppression. In a series of experiments, participants were required to shoot at targets. The ability to complete the task was measured by capturing the number of targets hit and shots fired. Experiments consisted of phases where fear-inducing threat cues were presented when the actions were performed. A significant reduction in the ability to complete the task was found when participants were shown these threat cues. This suppression in behavior by a Pavlovian threat cue is in line with findings where a general reduction in approach-related responses and an increase in freezing-like response in the presence of a threat [84, 85].

Unity Experiment Framework In [86] the *Unity Experiment Framework* (UXF) is discussed. UXF is a tool that allows behavior researchers to use the gaming engine Unity to design experiments for both VR and non-VR experiments. UXF simplifies the development of experiments and the measurement of data. It makes use of a structure typical to human behavior experiments, which is the *session-block-trial* model. Independent variables can be set to an experiment, session, block, or trial. UXF also allows for the continuous measurement of variables at the same rate as the display refresh rate. Finally, it has options for cloud-based experiments, making it possible to distribute the experiments to a larger group of participants.

Assessing Virtual Reality Various instruments exist to assess VR aspects. In [87] two presence questionnaires are discussed, which are reliable in the assessment of immersion in VEs. Both use a seven-point scale, with each item having opposing descriptors on both ends. The *Immersive Tendencies Questionnaire* (ITQ) is designed to be administered to participants before introducing them to the VE to measure possible individual differences in the abilities or tendencies of participants to immerse themselves in different environmental situations. The *Presence Questionnaire* (PQ) measures the degree to which a VE aspect generates a sense of presence for a participant, which should be used to measure individual differences in immersion in the VE.

Questionnaires also exist to assess sickness after the usage of VR. The *Simulator Sickness Questionnaire* (SSQ) could serve as a scoring system for the sickness induced by simulators and VE systems based on a list of symptoms [88, 89, 90]. A more recent questionnaire developed specifically for VR is the *Virtual Reality Sickness Questionnaire* (VRSQ), which is a motion sickness measurement index of a VE [90]. VRSQ is derived from SSQ but is a more efficient questionnaire with fewer questions, as some questions in SSQ were irrelevant for VR.

2.4.4 Combining Gaze Tracking & Virtual Reality

As mentioned earlier in Section 2.3.1, it can be analyzed what the participant is looking at in the virtual world when a VR solution is used with a built-in gaze tracker [12]. To do this, the 3D gaze vector going from the participant's eyes into the gaze direction needs to be calculated, which some eye-trackers can already determine, such as the HTC Vive Pro Eye [64]. Depth is theoretically possible but imprecise in most cases when the crossing point of the gaze from both eyes is used. VR has the advantages of having a 3D eye model and the complete knowledge of the distance between eyes and objects in the VE, which make it possible to calculate the depth of the gaze point in 3D space, assuming the spatial extent of the fixated object is large compared to the inaccuracies of the eye tracker. The 2D gaze location should be converted into a 3D gaze vector in the virtual world by normalizing the eye position of both eyes and then using the head position and orientation. The depth can be calculated based on the first intersection of the gaze vector with an object in the VE.

2.4.5 Challenges of Virtual Reality

VR has benefits in behavioral studies. However, it does pose some issues which might influence the behavior of the participants. These should be taken into account when designing VR experiments.

Motion Sickness A challenge in VR is the possibility of experiencing negative side effects, such as motion sickness and postural instability, which are induced by a *visual-vestibular conflict* (VVC) [91, 12]. In VR, this is caused by the eyes perceiving motion while the user is static or when latencies between head movement and VR display updates are present. In [91] it is shown that a mismatch between head motion and presented image in VR resulted in participants feeling motion sickness and postural instability. The background of the VR turned to a degree double that of the head movement on a vertical axis, which caused motion sickness in participants. In [92] the same effects were shown when the visual scene was delayed after the head movement in VR. In recent work done by [93] assumptions were made on what causes motion sickness in VR and how to reduce it. It is stated that motion sickness is induced by changing the head orientation without the user input, by modifying the field of view, by zooming in and out, by variations in acceleration, or by speeding up and down. It was also stated that motion sickness could be reduced, by maintaining the immersion from the start until the end of the experiment, by adding a frame reference within the peripheral vision that connects the user to the virtual world. Additionally, VR should not be treated as a camera by avoiding zooming in and out or using other camera tricks.

In [12] it is stated that dizziness and nausea are caused by latencies higher than 15-20ms, caused by the assumption that the VR world follows the same rules as the real world. It is also stated that a varying frame rate can also cause nausea, so developers should keep in mind the complexity of the experiment's VE. A solution to this is to design the VE in such a way the participant does not have to walk.

Motion sickness in VR also appears to be correlated to the sex of the participant [94]. It is found that females overall have a higher risk of experiencing motion sickness in VR than males. These results are backed by [95], which stated that females appear to be more subject to motion sickness in general.

Fatigue In [12] it is also discussed that using VR could cause fatigue. Keeping on the headset could also be exhausting since these devices could be a bit heavy and pull the head of the participant forwards, causing pressure/pain on the nose and neck. Additionally, wearing a VR device can be warm during summer days or in hot environments (as mentioned earlier in Section 2.2.3, where this can negatively influence facial temperature data). Another challenge is the disparity between vergence and focus. This disparity causes eye strain and limits the participant in extracting any depth information out of the focus of the lens (as mentioned earlier in Section 2.4.4). However, the latter is not a large issue for most people.

Eye Tracking in Virtual Reality Additionally, in [12] it is discussed that participants wearing glasses cannot always participate in VR experiments, as these do not allow the VR device to be placed on the participant or makes eye tracking inaccurate due to the lenses of the glasses. Eye trackers also need to be calibrated and validated, which requires extra time and can be disruptive during longer sessions. This also means that the participant cannot move the headset on the head anymore after calibration, so it must be sitting on the face comfortably at the beginning of the session. Fast movements in VR could lead to the headset moving on the face, which should be avoided. It also appears that some eyes are tracked easier than others (e.g., usually bright eyes are better tracked than dark eyes).

Behavior in Virtual Reality Screens have a higher effective resolution in the center compared to the outer parts, which influences the relationship between the head and eye movements since the participant needs to move the head to see an object of interest at a high resolution in VR [12]. Limited FOV also contributes to this effect. When in VR, the movement might not be natural as you have limited or no freedom to move around with most solutions. A solution is to eliminate the ability to move around and design the scene in such a way the participant does not have to move but can only look around.

Movement options exist, such as teleportation. Teleportation does not generate any optical flow, which reduces the risk of motion sickness [96, 97]. However, it breaks immersion since the user does perform an action that does not exist in real life. Additionally, it spatially disorients the participant. In [97] a teleportation technique called *Dash* is discussed. Dash quickly but continuously translates the user's viewpoint and generates a small number of optical flow cues, which appeared to improve spatial awareness compared to regular teleportation. However, because teleportation is not a natural way of movement, it is eliminated in most research [12]. Another solution is a tracking area, but for bigger VEs, this requires much space and a solution for cables or wireless transmission. An alternative is an omnidirectional treadmill, but this technology is not very well-developed and quite expensive at the time of writing.

Other issues are the limited depth information screens can convey, which makes it difficult to estimate the size of a virtual object [12]. Finally, interaction with objects in the VE requires hand tracking systems instead of the use of controllers.

3 Method

In this chapter, the methodology of the experiment will be discussed. In Section 3.1 the characteristics of the recruited participants will be discussed. In Section 3.2 the required materials, such as the hardware devices used to record data and to run the experiments will be discussed. In Section 3.3 the used stimuli will be discussed. Section 3.4 will discuss the VE, in which the participants can look around and in which the stimuli are presented. Section 3.5 will explain the experimental procedure. Section 3.6 will explain the steps taken for pre-processing the recorded data and will define measurements used for the statistical analyses.

3.1 Participants

The participants were required to be fluent in English and to be at least 18 years old. They were recruited in the vicinity of the Utrecht University and via the university's participant recruitment group on Facebook. Therefore, this group consisted to a large extent of students. All participants received a gift voucher for their participation and provided written informed consent.

3.1.1 Health Requirements

The participants were required to have unimpaired vision. For instance, they were required to not suffer from any issues which largely impact their natural viewing behavior, such as blindness or other damage to the eye. Additionally, participants were not allowed to wear glasses during the experiment. This was impossible when wearing the VR device and could negatively influence the recorded data from the eye tracker due to the refraction of light through the glasses. Exceptions were made for participants who wore lenses most part of the day under the condition they regard this as their natural vision. In these cases, the participants were asked to keep their lenses in during the session to make sure they would not show adjusted or unnatural gaze behavior. Further health requirements of the participants were not suffering from motion sickness, nausea, and epilepsy. Lastly, the participants were required to be able to stand up for 30 minutes.

3.2 Materials

The experiment was conducted at the *Human-Centered Computing Lab* at the Utrecht University's Buys-Ballot Building. The setup used to conduct the sessions can be seen in Figure 3. This setup consists of various hardware devices, which gather behavior data or present the stimuli. These devices will now be discussed in more detail.

3.2.1 Wii Balance Board

To measure body sway, participants stood on a Nintendo Wii Balance Board. This device could be connected to a PC via Bluetooth. This device has a sampling frequency of approximately 40 Hz. It has four pressure sensors located at the corners of the board, which measure the pressure in kilograms. These can be used to track the COP in the anterior-posterior (AP) direction (y-direction) and the mediolateral (ML) direction (x-direction). These directions are illustrated on a Wii Balance Board in Figure 4.

FIGURE 3: The experiment setup used to gather behavior data. The participant takes place on the Wii balance board and puts on the HTC Vive Pro Eye. On the left and right two of the three lighthouses can be seen, which are mounted on tripods. These track the VR headset's position and orientation.

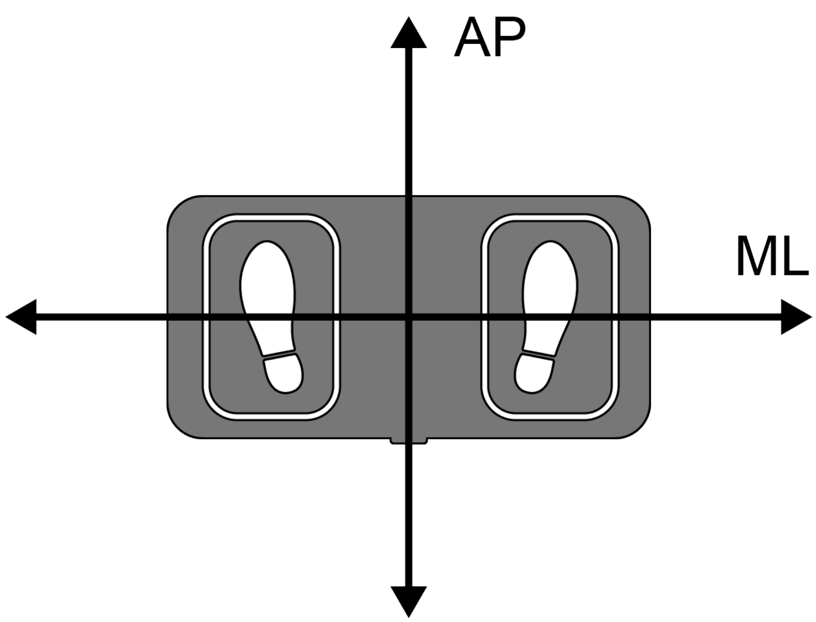


FIGURE 4: Illustration of the anterior-posterior (AP) and mediolateral (ML) directions on a Wii Balance Board from a top down perspective.

3.2.2 HTC Vive Pro Eye

Stimuli were presented in VR using an HTC Vive Pro Eye VR headset. This device has dual OLED 3.5" displays with a resolution of 1440 x 1600 pixels per eye, a refresh rate of 90 Hz, and a total FOV of 110°. This VR headset makes use of base stations, also called lighthouses, for outside-in tracking. This technique allows for high accuracy mirroring of head movements in the VE, which has already been discussed in Section 2.4.1. Although two base stations are sufficient, three were available and were used to track the VR headset, which would minimize the possibility of incorrect tracking or loss of signal with the VR headset. These were mounted on tripods, which were placed in a triangle shape around the participant to have a large tracking coverage of the headset from multiple sides. Additionally, the headset has capabilities to track head movement and orientation of the participant over time, but since similar patterns of movement are expected to be also seen in the balance board data, this data will not be used in this research.

3.2.3 Tobii Eye Tracker

Gaze in VR was analyzed using the HTC Vive Pro Eye's built-in Tobii eye tracker. The tracker has a sampling frequency of up to 120 Hz, covers the whole FOV of 110°, and has an accuracy between 0.5° and 1.1° within 20° from the center of the lenses. It can easily be set up for each participant using a five-point calibration system. This tracker records data about each individual eye, such as gaze origin, gaze direction, normalized pupil position in the sensor, pupil size (mm), and eye openness. Additionally, it makes use of Unity's ray-casting system to determine the focus objects from the gaze vectors provided by the gaze tracker. This system traces a ray or vector and provides information about the first object intersecting the ray, such as intersection point and what object has intersected the ray.

3.2.4 Experiment Software

The experiment software connects to all hardware devices, handles data collection, and presents the stimuli to the participants. This software was custom developed for this research using Unity engine version 2020.2.6f1. The *Universal Render Pipeline* (URP) template settings were picked for the project, as it allows for easy adjusting of render settings for running on VR platforms. The VR platform used here was the SteamVR platform, which contained many already implemented features for interacting with the VR headset, which required minimal coding and sped up the development process significantly. The UFX Unity package (Section 2.4.3) was used to build the skeleton of the experiment software. This package provided many features for setting up trials and for handling the recording of data. BitLegit's *WiiBuddy* Unity package [98] was used to obtain raw Wii Balance Board input from its Bluetooth signal.

Graphical User Interface The experiment software provides an easy-to-use graphical user interface (GUI). This GUI allowed the experimenter to enter a participant ID and to configure session parameters before starting. After configuration, the experimenter could launch the Vive Pro's eye calibration, start the session, or log trials that would be invalid using the GUI's buttons. Furthermore, this GUI features insights into if all hardware is functioning and displays the measured sensor data. This GUI can be seen in Figure 5.

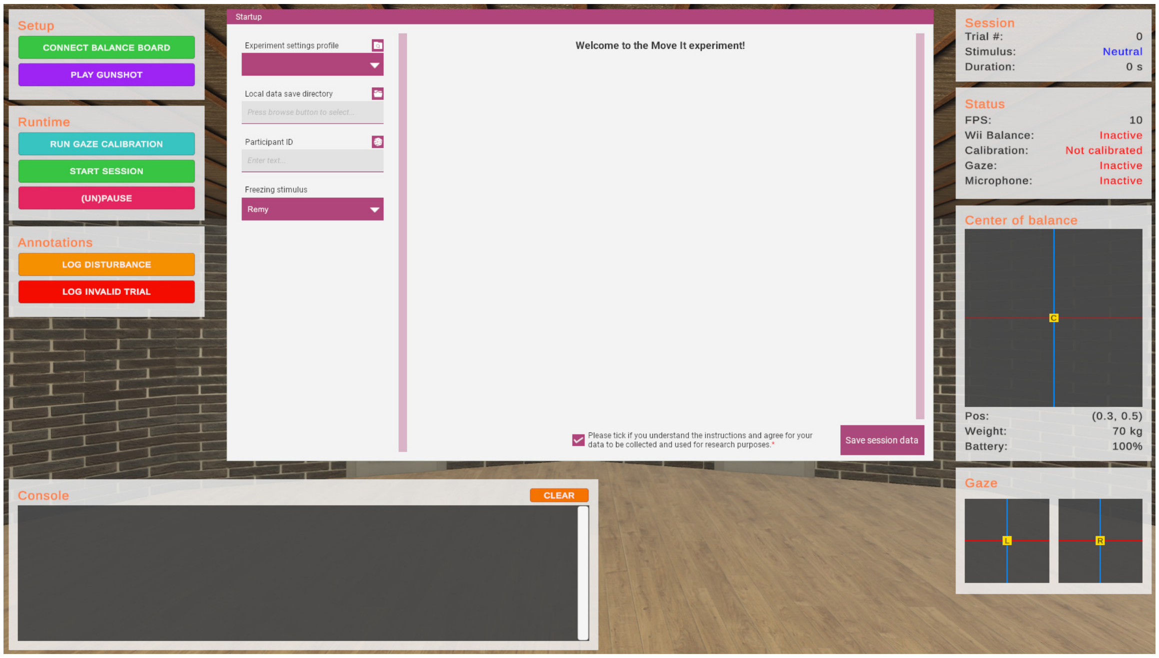


FIGURE 5: The GUI of the experiment software. The top left panels contain buttons trigger various functionalities, such as connecting to the hardware, running calibration, and starting the session. The center panel is used to configure the session. Here you can provide the participant ID, data storage directory, and other session parameters. The right panels inform the researcher about session progress, performance, hardware status, and insights into the last read sensor data. At the bottom, there is a console that logs all events with a timestamp during the session.

3.2.5 Desktop PC

The experiment software was run on a Windows 10 machine. This machine has an Intel Core i7-6700K quad-core CPU (base speed 4.0 GHz, but overclocked to run at 4.4 GHz), an MSI GeForce (NVIDIA) RTX 3080 Gaming X Trio 10G GPU (base clock of 1.44 GHz; maximum turbo boost clock of 1.815 GHz; 8,704 cores; 10 GB of GDDR6X VRAM), and 64 GB of DDR4 RAM (running at 2667 MHz). This configuration was by no means a representation of how much the workload was to run the experiment software itself, but this was the hardware available for this experiment. However, this hardware allowed for running the session at a frequency higher than 90 Hz almost at all times, with minimal drops to lower frequencies (still within the 80-90 Hz range). This means software latency was minimal such that it reduces the chances of inducing motion sickness and nausea, which should mitigate some of the challenges described in Section 2.4.5. Additionally, this would make sure the hardware would provide data at the highest frequency possible. Every cycle of the software, the hardware would be polled for changes in the last read sensor values. This means that the amount of detail of the data relied on the temporal frequency at which the software ran (except for the balance board, which ran on a much lower maximum frequency). Due to this reliance on the software's run-time frequency, temporal dips in the software frequency also resulted in temporal reductions of detail in the data signal, which means that the amount of detail in the data could also fluctuate over time.

3.2.6 Questionnaires

Before the VR session, a questionnaire had to be filled in. This questionnaire consisted of the *State-Trait Anxiety Inventory* (STAI), the *Tonic Immobility Scale* (TIS), and various health- and vision-related questions. The first 13 questions were concerned with the participant's age, sex, prior VR experience, and health. The STAI consists of 40 questions in total based on a four-point Likert scale. The STAI-state consists of the first 20 questions, which measure anxiety about a specific event. For this research, it was used to assess anxiety at the moment of filling in the questionnaire before the VR session, to see if participants might show adjusted behavior. The STAI-trait consists of the remaining 20 questions and assesses the anxiety level as a personal characteristic. For both sets of questions, a final score will be calculated within the range of 20-80, where a higher total score means a higher anxiety intensity. The TIS consists of ten questions on a zero to six-point scale, which measure involuntary fear behavior, such as freezing and physical immobility, during a fear-inducing event the participant was instructed to think about. Three of the questions are used to calculate a total score for fear within the range of 0-18. The remaining seven are used to calculate a total score for immobility within the range of 0-42. Additionally, a total score is calculated for all TIS questions combined within the range 0-60. A higher total score means a higher intensity of involuntary fear behavior. After the VR session, one final questionnaire had to be filled in. This questionnaire was concerned with how the participant experienced the VR sessions. It contained 12 questions, which were based on a five-point Likert scale. They were concerned with assessing to what extent the participant felt immersed, how realistic the environment appeared to them, if the session had made the participant feel sick, and if they experienced fear or froze up when approached by the two characters.

3.3 Stimuli

The stimuli were obtained from Adobe's *Mixamo* website, which contains free-to-use, rigged, and animated 3D characters. From this website, the characters *Remy* and *Josh* were chosen to be used as stimuli during the experiment. These characters can be seen in Figure 6. The website contains mostly fantasy characters, such as monsters, but these characters appeared most human-like and had the most casual clothing. Additionally, these characters were picked as they had identical skin colors but were different enough to distinguish from one another. The characters were programmed to look the participant into the eyes during each trial.

3.3.1 Stimuli Variants

Both characters had a neutral and fear-inducing variant. The fear-inducing variant's face of each character would be adjusted to look angry, as angry faces were used as fear-inducing stimuli in previous works, as described in Section 2.2.1. To appear angry, the corresponding anger emotion facial AUs were activated, as listed by FACS [11]. These AUs are listed in Table 4. Additionally, the fear-inducing variants would hold a rifle and point it at the participant around chest-level height. An example of Josh's fear-inducing variant can be seen in Figures 7 and 8.

3.4 Virtual Environment

The VE used for the research was hand-made with Unity's *ProBuilder* tool, which is a tool for designing and prototyping 3D environments rapidly. Two rooms were designed to resemble each other as much as possible.

FIGURE 6: The character models of *Remy* (left) and *Josh* (right)

FIGURE 7: The characters from the point-of-view of the participant after entering the room. On the left Remy is the neutral stimulus and on the right Josh is the fear-inducing stimulus. Josh has an angry facial expression and holds a rifle, which is aimed at the participant.

FIGURE 8: A different perspective of Josh as a fear-inducing stimulus inside the Unity editor.

AU Identifier	Facial Expression Description
4	Eyebrows are drawn/pulled together
5	Raised upper eyelids
7	Tensed lower eyelids
23	Lips tightener

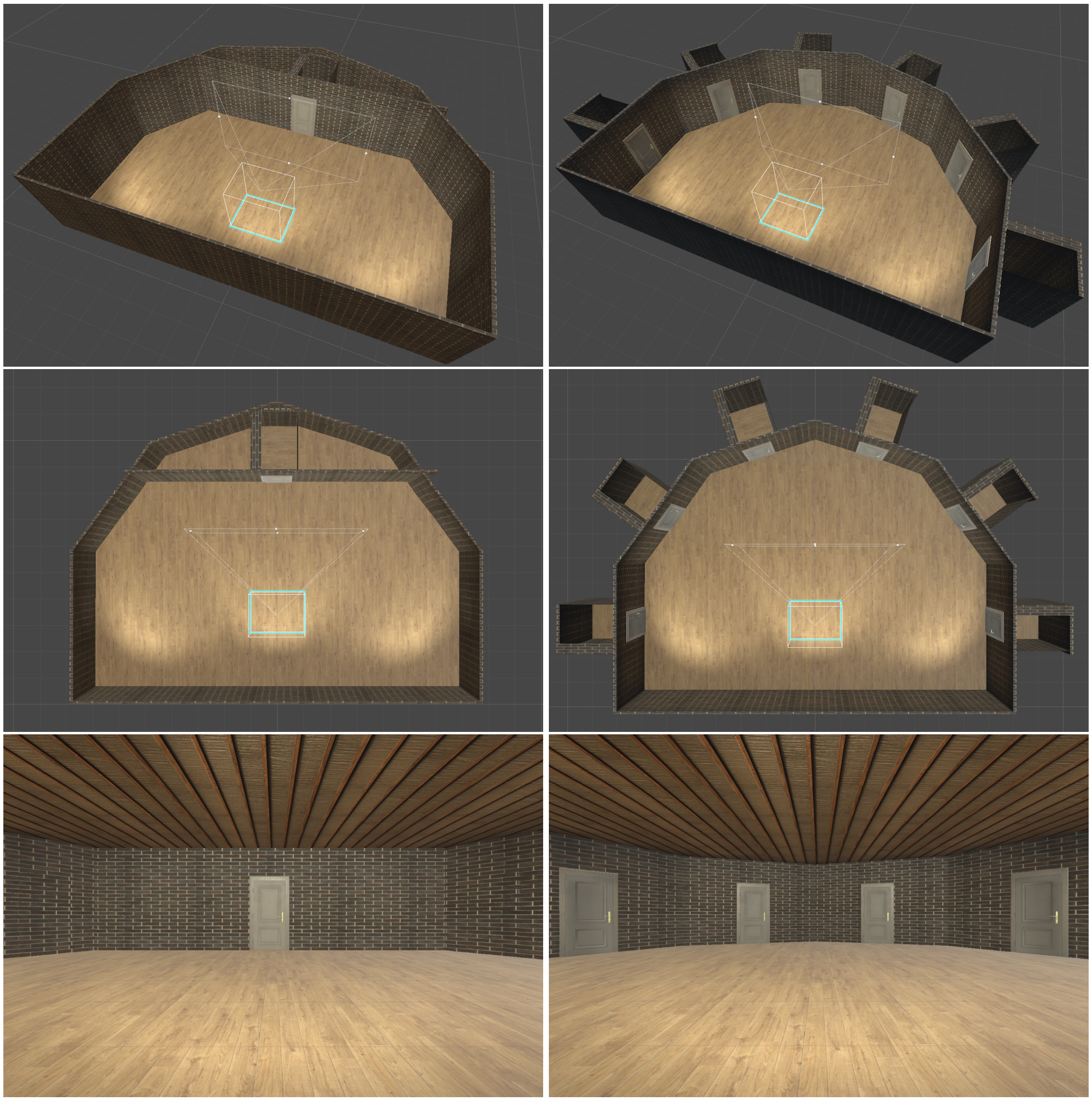
TABLE 4: The facial expressions of anger and their corresponding action units as described by FACS.

These can be seen in Figure 9. One of the rooms featured a single door, and the other room featured six doors. The participants were placed with their back facing the back wall, looking into the room. The room is designed in a way that as many doors possible are within the FOV of the participant, without requiring too many body movements to turn towards doors on the left or right, as these movements might pollute the captured body sway data. The doors of these environments will open during the trials. An audio clip of the door opening will be played through the participant's surround headphones to make sure the participant is always quickly able to find an opening door, even when it is located outside of their FOV.

The environments were textured and lighted in a way it would look sufficiently realistic and not distracting. Various lights were placed, which would cast shadows in the room, to make the elements appear less flat.

3.5 Session Procedure

An overview of a complete experiment session is given in Table 6. A session would take approximately 45 minutes in total. The participant would wait outside the lab until the researcher let the participant inside,



(A) The single door environment

(B) The six doors environment

FIGURE 9: Side-by-side comparisons of the two VEs from multiple perspectives. The blue rectangle indicates the participant's location in the VE.

to prevent disturbing any ongoing sessions. Each session segment will now be explained in more detail by chronological order of execution.

3.5.1 Briefing

Participants were given a briefing at the start of the session. During this briefing, they were told how the session would be structured. Additionally, the participant was instructed how to take place on the Wii Balance Board and how to put on the VR headset for when the VR segment would commence. The participant was instructed to keep both feet on the board and to keep the hands out of the pockets. The participant was further instructed how this VR segment would be organized and that two characters would make their appearance in the VE, and that the participant would be placed in a room with a single or multiple doors.

After this briefing, participants had to fill in a consent form for recording sensor data, as well as fill in the first questionnaire. This part would approximately take 10 minutes.

3.5.2 Virtual Reality Segment

During the VR segment, the participant had to take place on the balance board and had to put on the VR device. Once standing comfortably, they would give a vocal signal they were ready to start.

Eye Calibration At first, the eye tracker had to be calibrated, by using the Vive's integrated calibration system. During this calibration process, the participant had to adjust the position of the device on the face and had to adjust the interpupillary distance (IPD) using a wheel that was located at the front of the device. This was done to accurately track the pupils and to optimize the display quality, as the IPD is unique for each participant. If this is not configured properly, the display would be unsharp. After this step, the participant was asked to focus on a dot, which would appear in five different locations (one at the center, and the others in each display corner).

Introduction Upon completing the calibration, the participant was asked to give a final vocal signal to start. From this point on, the participant and researcher were not allowed to communicate until after the VR segment. The VR session would start with a 60 seconds long introduction, to make the participant get used to the VR environment and the stimuli. During this segment, the participant is standing in a room with a single door. In this room, both stimuli were shown as their neutral variant (neutral facial expression, not holding any weapons). They would walk around in the room in circles, not looking the participant right into the eyes.

Trial Structure After the introduction, the trials would commence. The trials were structured similarly throughout the remainder of the session but with small differences between the blocks, which will be described in more detail later.

The general chronological structure of a single trial can be seen in Table 5 and an example trial is visualized in Figure 10. The mean duration of a single trial was 22.65 s (σ = 2.00 s). The mean total duration of the VR segment (excluding the eye calibration) was 1269.94 s (σ = 9.64 s), which translates to just above 21 minutes on average.

The duration of some of the events has been pseudorandomly generated by the *Xorshift 128* pseudorandom number generator, which is an efficient generator that requires limited coding to implement [99]. This generator was already implemented by Unity. Randomness is added to prevent trials from becoming too repetitive and to add some uncertainty.

During each trial, either the fear-inducing or the neutral stimulus would enter through the door. The order of the presentation of the stimuli is chosen, again pseudorandomly, under the condition that the same stimulus will not be shown for more than three consecutive trials. This is to make sure both stimuli will be shown during all the blocks. In a worst-case scenario, one of the stimuli might not be shown at all without this condition, given the (pseudo)random nature of them appearing. As can be seen in Table 5, the trial might take slightly longer in case of a fear-inducing stimulus, as extra time should be taken into account for the rifle shot taking place. During this rifle shot, a loud sound effect will be played and a muzzle flash can be seen.

Description trial event	Event duration (s)
Screen fades from gray to clear, the room is now visible	1
Short period where nothing happens to let eyes adjust to lighting conditions	4
The door opens	0.5
Stimulus enters the room and stops at 2 m distance away from the participant	5-6
A random idle time where the stimulus stands still	2-5
In case of a fear-inducing stimulus: character will fire rifle at participant	1
A short delay	2
The screen fades to gray	1
The screen will stay gray for a random duration	4-7
	Approx. 23 sec in total

TABLE 5: The general chronological structure of a single trial

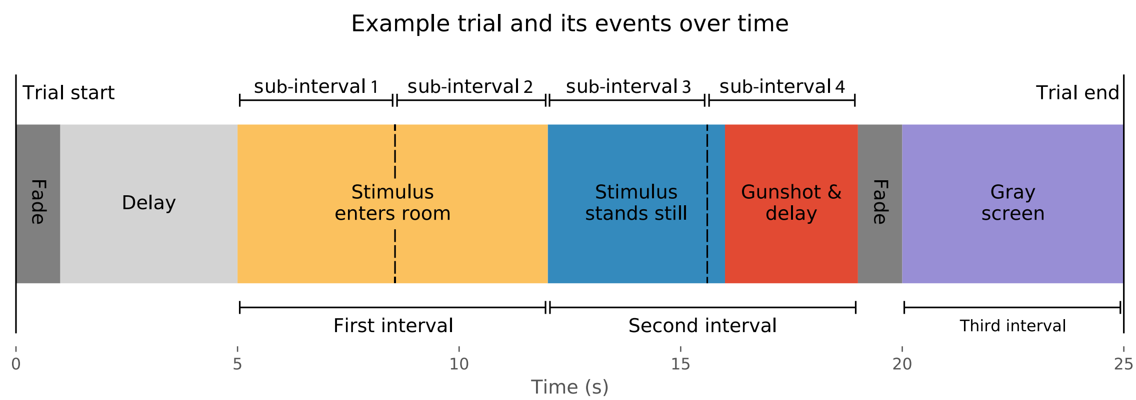


FIGURE 10: An example trial timeline. The three intervals which remain after trial segmentation are visualized in colors. The parts which are filtered away are visualized in shades of gray. For reasons later described in this thesis, the trial intervals could also be split up at the middle into sub-intervals.

After a short time, the screen fades to gray and will stay on gray for a short time, until the next trial commences. These serve as a transition between trials and allow the VE to be reset (close door(s), reset characters, etc.) without distracting the participant. The color gray was chosen, as a black or a white fade could trigger the pupillary light response. This response will dilate the pupils in case of darkness and constrict the pupils in case of brightness [100]. The color gray lays between black and white when looking at lightness intensity, thus hopefully influencing the pupil sizes the least. Additionally, a delay of four seconds is held after when the screen fades back from gray to clear at the start of the next trial, to make sure the eyes get used to the lighting conditions of the VE and the pupil size stabilizes. These measures are taken to reduce any unwanted pupil size behavior, which is not caused by any of the stimuli.

Single Door Blocks After the introduction, the habituation block would start. This block consists of four trials, where during each trial one of the two characters would enter the room through the only door in the environment. This block does not contain any fear-inducing stimuli (again, no angry faces or rifles) yet, as its purpose is to make the participant aware of what happens when the stimuli enter through the door. The two characters alternate during this block, to make sure both characters were both seen twice. In this block, the order is not randomly generated, as opposed to the other blocks.

Now that the participant has experienced and has gotten used to the neutral characters entering the room, a new block will start. This block will contain fear-inducing stimuli and consists of eight trials, where one of the characters has been given the role of fear-inducing stimulus. From this block on, this character will enter the room pointing a rifle at the participant and will have an angry facial expression. The character that is chosen to be the fear-inducing stimulus is alternated per participant, so the quality of the stimuli could be assessed afterward. From now on, this block will be referred to as *block one*, as it is the first block containing fear-inducing stimuli.

Questions Between Blocks Now that the first fear-inducing stimulus block has been completed, the participant will be asked to answer two questions by voice. These questions appear after every block containing fear-inducing stimuli during the VR segment, and the two questions asked will always be the same. This is the only moment the participant is allowed to speak during the VR segment. However, the participant still had to keep on the VR device to prevent breaking the immersion with the VE. These questions consisted of giving a rating from a 0-5 point scale to what extent they were afraid of both the fear-inducing and neutral character. In total, these questions would be asked four times during the VR session.

Six Door Environment Blocks After the single door blocks, the environment will switch over to a nearly identical room, which now has six doors instead of one (Figure 9). Two blocks consisting of eight trials each will now take place. During these trials, a single stimulus will enter through these doors. These blocks will now be referred to as *block two* and *block three*, respectively.

During block two, the stimuli will enter through fixed doors. The fear-inducing stimulus will always enter through the second door from the left, and the neutral stimulus will always enter through the second door from the right. After this block, the two questions will be asked for the second time.

After answering the two questions, block three proceeds with the stimuli now enter through (pseudo)randomly chosen doors, instead of fixed doors, which is the only change compared to block two. Once this block has been completed, the two questions will be asked for the third time.

One final block of eight trials will take place in this environment, which will now be referred to as *block four*. Now, the two stimuli will enter simultaneously. Two adjacent doors are chosen (pseudo)randomly, to make sure both stimuli are within the FOV of the participant. Both stimuli will be assigned one of these doors (pseudo)randomly. Another adjustment in this block is that the fear-inducing stimulus will not fire its rifle. However, it will still aim the weapon at the participant. One final fourth time the two questions will be asked.

Extinction Block The environment will now switch back to the environment with a single door. This final block will consist of ten trials and will present the stimuli again one at a time. Both stimuli will be presented in their neutral form, so the fear-inducing character will not hold a weapon here or have an angry facial expression. This block serves to remove the association of fear or danger with the fear-inducing stimulus. Once this block has finished, the participant is instructed to take off the VR headset and to step off the balance board.

Session segment	**Number of trials**	**Segment duration (min)**
Briefing, questionnaires	×	10
Calibration gaze tracker	×	1
VR introduction	×	1
Habituation block	4	2
(block one) Single door block	8	4
(block two) Six doors block, fixed stimuli doors	8	4
(block three) Six doors block, random stimuli doors	8	4
(block four) Six doors block, two stimuli simultaneously	8	4
Extinction block	10	6
Questionnaires, debriefing	×	5
		Approx. 45 min in total

TABLE 6: The general chronological structure of a single session. The blocks in which the sensor data will be analyzed for fear behavior are presented in italics. All times are approximates, due to the pseudo-random duration of trials.

3.5.3 Wrap Up

After the VR segment, the final questionnaire was filled in, which was concerned with the participant's well-being after the session and the VR experience. Upon completion, the participant received a gift voucher for their participation.

3.6 Data Analysis

In the following section, the steps taken to analyze the recorded data are discussed. For the sensor data, the cleaning process will be discussed first, as some of the raw sensor signals contained missing values or noise or contain less informative data. After the cleaning steps, the data is split into intervals based on their timestamps. This makes it easier to compare the data at different intervals throughout the trials. Finally, the measurements used to compare data will be discussed.

3.6.1 Trials of Importance

Only the trial sensor data from the four blocks represented in italics from Table 6 will be used for the data analyses. The habituation block is used to make the participant familiar with the general structure of a trial. The extinction block is meant to undo the fear associated with the armed character. Therefore, they are not useful for detecting fear behavior, and their data is discarded in the analyses.

Additionally, the data gathered during the first trial of block two is filtered away, as it had incorrect time labels due to a software bug that went unnoticed until after the sessions.

3.6.2 Trial Segmentation

Sensor data is captured throughout the entire duration of each trial, and every data entry has a timestamp. Based on these timestamps, the recorded data will be labeled with the trial events listed in Table 5. This is done to make it easier to filter the sensor data based on trial events. The sensor data at less important trial events is discarded, which is data at the moments between the start of the trial until the moment the door opens and the moments on which the screen fades to gray. The moments before the door opening contain no stimuli, so they do not provide useful data. The time in which the screen fades to gray could temporarily adjust the participant's behavior, especially the gaze behavior due to changes in lighting conditions. Therefore it is also filtered away.

Intervals Per trial, three intervals remain. These consist of the segment where the door opens and the character enters the room, the segment where the stimulus stands still in front of the participant (and in case of a fear-inducing stimulus, fires the rifle), and a segment where the screen is gray. Note that these segments differ in duration per trial due to the added pseudo-randomness discussed earlier. These intervals are visualized in Figure 10.

Sub-Intervals In some cases, the intervals will be split up from the middle into two sub-intervals each. Splitting them up resulted in multiple smaller intervals, which means more focused data was available over time. For instance, when looking at the first interval, it is only known how the participants behaved in general throughout the period the stimulus entered the room. By splitting the interval up, the behavior could also be analyzed when the stimulus entered the room and was still far away and when the stimulus walked closer towards the participant. By splitting up the intervals, it was possible to analyze further how behavior was adjusted throughout each trial. In the results of Chapter 4, this will be done when a found pattern in an interval requires closer inspection or if it is suspected that a pattern occurs in a shorter period. Additionally, the sub-intervals will be used to determine whether found behavior occurs with a delay. An visualization of such sub-intervals of an example trial can be seen in Figure 10.

3.6.3 Measurements

The following section discusses the measurements used to analyze differences in behavior. These measurements are calculated for each trial (sub-)interval and for each participant.

Body Sway Previous works analyze body sway by looking in both AP- and ML-directions [2, 3]. To keep in line with these works, for both the AP and ML direction components of the COB the absolute difference between two subsequent observations (1D location of the COB on the AP or ML axis for a moment in time) p and q was used as a measurement for body sway:

$$d(p, q) = |q - p| \tag{1}$$

Additionally, the 2D Euclidean distance between two subsequent observations (2D location of the COB at a moment in time) $\mathbf{p}$ and $\mathbf{q}$ was used as a measurement for body sway:

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{(q_{\mathrm{AP}} - p_{\mathrm{AP}})^2 + (q_{\mathrm{ML}} - p_{\mathrm{ML}})^2} \tag{2}$$

These would be calculated for each timestamp of a trial. Per trial interval, the mean would be calculated for these body sway measurements, as each trial (interval) has a different duration, and comparing the measurements would be more difficult. This results in nine distance measures per trial (three distance measures times three trial intervals).

Gaze Fixations Defining measurements for gaze fixations was less straightforward, as the gaze tracker provided limited insight on the calibration performance (only a success or failure message). Therefore, gaze tracking accuracy could not be statistically validated. Due to this, it would be hard to determine whether short, quick movements in the signal were caused by the eye moving in saccades or sensor inaccuracies. Therefore, measuring fixations was done by taking the fraction of the time t_c spent looking at a specific object of interest c in the VE over the total duration of the trial segment t_{seg}.

$$P(\text{c}) = \frac{t_{\text{c}}}{t_{\text{seg}}} \tag{3}$$

Per object, this results in a value between zero and one. The objects of interest are objects in the VE that contain a collision bounding box. These objects are the body parts of the character models, the rifle, and the environmental objects (walls, doors, ceiling, etc.). For both eye locations in the VE, a ray is cast in the direction the eye is looking, which represents the gaze vector. The first object containing a collider that intersects with the eye's gaze vector is seen as the focus object of that eye, as discussed earlier in Section 3.2.3.

Additionally, the gaze tracking software provides combined eye data by combining the data gathered from both individual eyes. This results in a third gaze ray (and can be seen as a virtual third eye), which has its origin located between the two eye origins and points in the direction of the intersection of both eye's gaze rays, which can be seen as the gaze target or focus of both eyes combined. The gaze rays are visualized in Figure 11. Instead of looking at the focus objects by looking at the individual eyes, the focus object of the combined eye was used. The benefit of using this combined eye is that on some occasions (such as looking at the edge of an object), one of the eye rays might intersect with a different object. In such cases, the combined eye takes away the decision of which eye's focus object to use. This combined eye handles signal loss better (for instance, caused by blinking). In case of signal loss in a single eye, it will pick the focus object from the other eye which still has a signal. This should reduce the number of missing observations. The downside of this approach is its reliance on both eyes to be calibrated properly. If one of the eyes is inaccurately tracked, the data returned by the combined eye automatically also is of lower quality.

The gaze fixations were calculated for (groups of) objects. For instance, this is calculated for objects such as the head, the weapon, or the full body of each stimulus. Angry faces induce fear in humans, as discussed in Section 2.2.4. Therefore, analyzing the time spent looking at the head of the fear-inducing stimulus might give more insight into fear behavior. This will be discussed in more detail later in Section 4.4.1.

An attempt was made at assessing the quality of the gaze tracker and the gaze fixation signals. Two spheres were placed in the VE roughly at distances 2-3 m (representing a stimulus standing still in front of the participant) and 5-6 m (representing a stimulus entering through a door) away at eye height, which would resemble the heads of the characters. The researcher would fixate on each of the spheres for approximately a minute and later study the signals of both eyes and the combined eye globally for incorrect gaze targets. When fixating on the sphere close by, the signal gave steady results and minimal errors in gaze targets for both eyes (therefore, also the combined eye). The sphere placed farther away also returned a relatively steady signal with a few more occasions where the gaze targets were incorrect. Especially when fixating at the edges of the sphere, it would sometimes lead to incorrect gaze targets in the signal per eye. The combined eye returned the least amount of errors in gaze targets, so it was concluded that the combined eye signal was suitable for this research's purpose.

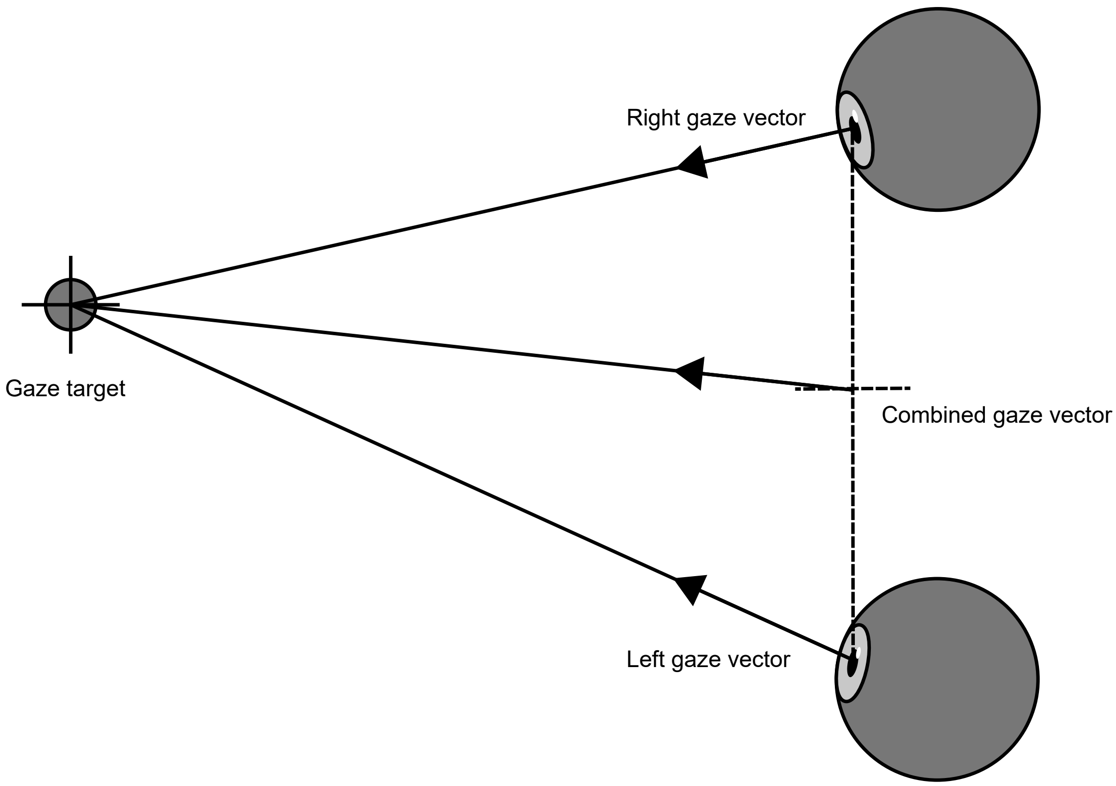


FIGURE 11: A visualization of the gaze vectors provided by the gaze tracker. The combined eye's gaze vector is also visualized and has its origin between the two regular eyes.

Pupil Size Previous works claim that the pupil dilates when participants were exposed to arousing or negative content, as discussed in Section 2.2.3. The pupil size in millimeters is recorded over time. Per interval, the mean pupil size is calculated.

3.6.4 Filtering Away Third Interval

The gaze data was filtered away during the third interval, as the participant could not see anything due to the screen being gray. Therefore, this data did contain incorrect gaze targets. The balance sensor data at the end of the trial could still be relevant to detect changes in movement behavior after a stimulus has been presented. To determine whether this data was informative, the mean values of the Euclidean distance measurement per stimulus type were analyzed. This was done for the trials of block one. This block would not require extra movements from the participant to locate the door or stimuli, as these will always be directly in front of the participant. This reduces the amount of movement in the data not caused by the stimuli, making it the most "clean" balance data for this analysis.

The mean values of the Euclidean distance measurement per interval for and stimulus type are visualized in Figure 12. In this figure, it can be seen that for the last interval, the mean values are almost equal when comparing the stimulus types, which is not the case for the first two intervals. From this visual inspection, it was concluded that this interval was not informative enough for balance data as well, meaning further analyses will focus only on the first two intervals.

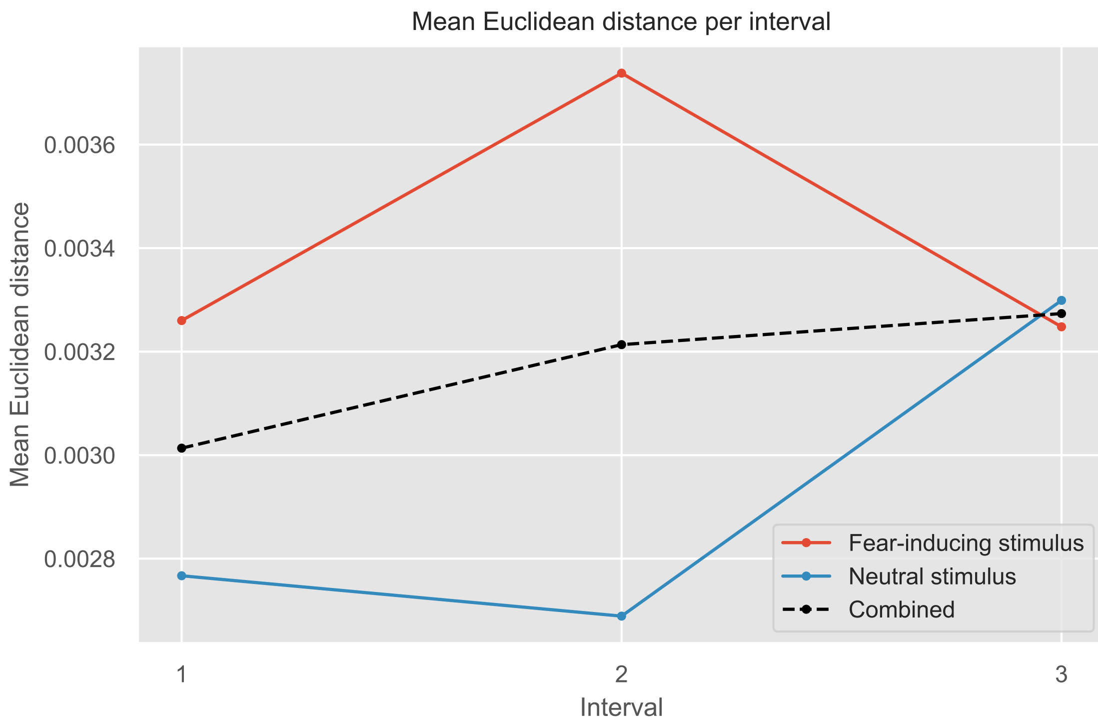


FIGURE 12: Plots of the mean Euclidean distance per interval and for both stimulus types.

3.6.5 Pre-Processing of Gaze Signals

Overall, the recorded data required minimal pre-processing. However, the gaze tracker's data did require some pre-processing, as it sometimes lost the pupils either by failed tracking or during blinks.

The pupil size's signal contained missing values and some signal noise, which could have been caused by inaccuracies of the gaze tracker. Therefore, the raw signal was first pre-processed with a forward filling step, which replaces missing values with the last read value of the sensor. This was done as the signals did not seem to have large jumps in the values and the intervals with missing values were of relatively short duration (less than half a second). To filter away the remaining noise and to smooth out the forward-filled signal values, a rolling average filter with a kernel size of 20 observations (which translates to roughly 22 ms, given a recording frequency of 90 Hz). This kernel size was chosen empirically and gave the best results. A sample pupil size signal before and after pre-processing is visualized in Figure 13.

It was challenging to determine if the participant was looking at an object or was blinking when the gaze signal was lost. Therefore, the decision was made to not fill in missing focus objects in the gaze target signals but to filter these observations away during analyses instead and to assign them the *Nothing* target in plots. No further signal processing steps were taken, as fluctuations in the gaze targets might also have been caused by saccades or inaccuracies of the sensor. This was hard to be determined, as there was limited insight into the gaze tracking device's accuracy.

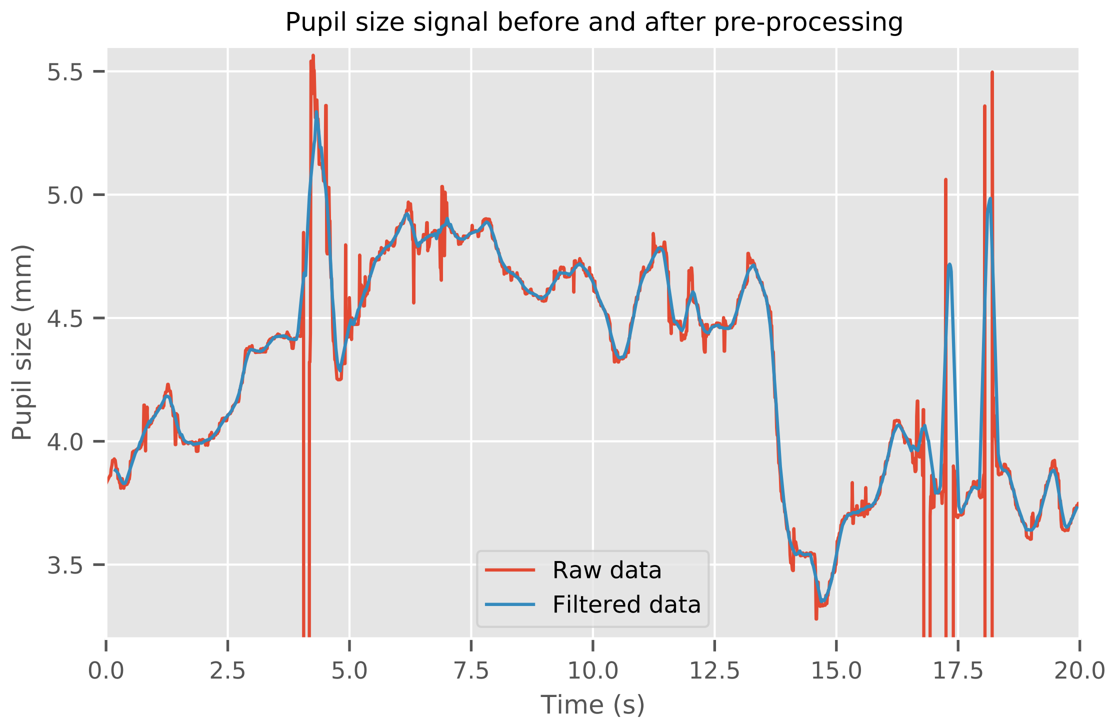


FIGURE 13: Example pupil size signal before and after pre-processing. It can be seen that sharp swings when the signal is lost (and it shoots to -1) are to a large extent reduced while maintaining the signal's shape (e.g., in the range of 3-5 seconds). It can also be seen that subtle noise has been reduced (e.g., in the range of 5-15 seconds). However, it does seem to struggle on some occasions (e.g., around 17 seconds).

3.7 Statistical Methods

The measurements described in Section 3.6.3 will be analyzed statistically using repeated measures analyses of variance (ANOVA). The answers to the questionnaires will be analyzed using one- or two-way ANOVA. For all statistical analyses using ANOVA, alpha was set at .05. For the body sway and gaze fixations measurements, (sub-)interval, block, and stimulus type (neutral, fear-inducing) are used as within-subjects factors. For the pupil size measurement, the eye (left, right) will also be included as a within-subjects factor. For the questionnaire data, the within-subjects factor were gender (male, female), prior VR experience (yes, no), stimulus type (neutral, fear-inducing), or character model (Remy, Josh). When Maulchy's Sphericity is not assumed, the degrees of freedom will be corrected.

3.8 Visual Inspections

Further inspections on the sensor data will be done on a participant level, as it could occur that participants do not all show similar behavior patterns. Additionally, it could occur that individuals show short-lived behavior patterns, which were not found in the general statistical analyses. Therefore, further analyses will be done based on visualizations of the measurements per participant or the measurements over time.

4 Results & Discussion

This chapter will show and discuss the results extracted from the data gathered during the experiments. Section 4.1 will first discuss all self-reported data, which is gathered from the questionnaires before and after the VR session. Section 4.3 will discuss the results gathered from the body sway measurements. Section 4.4 will discuss the gaze behavior during blocks one, two, and three, which featured one type of stimuli during the trials. Section 4.5 will discuss the gaze behavior during block four, which featured both types of stimuli simultaneously during the trials. Finally, Section 4.6 will discuss whether the found effects manifested with a delay.

4.1 Questionnaire Data

The following section will discuss data gathered from the questionnaires before and after the VR segment. This data will give an overview of the participants and how they experienced VR. Additionally, it is used to assess the overall quality of the stimuli. This data will further be used when analyzing the sensor data.

4.1.1 Participant Demographics

Nineteen participants (mean age = 24.63, σ = 5.92, min age = 19, max age = 42) were recruited for this study, of which 57.9% was male, and 42.1% was female. Thirteen participants (68.4% of total) claimed to have prior VR experience in any form.

Well-Being of Participants The first few questions were concerned with how the participant felt before the VR session. All participants claimed not to feel nauseous. Four participants (21.1% of total) claimed to suffer from motion sickness to some extent. The questionnaire after the VR session would eventually determine if participants suffered from motion sickness during the VR session. Therefore, these results only indicate the likeliness of motion sickness after the VR session and to what extent this was caused by the VR session. One of the participants claimed to feel fatigued before the VR session. None of the participants claimed to suffer from epilepsy. From the results could be concluded that the participants were healthy and capable enough of running the VR session based on their well-being.

Vision of Participants None of the participants claimed to wear any glasses daily or to suffer from (partial) blindness. Two participants wore lenses daily and they were instructed to wear their lenses during the VR session to make sure they would not show adjusted viewing behavior. One participant claimed to squint, which resulted in the inability to succeed gaze calibration during the session, thus inaccurate gaze data was recorded, so this participant's gaze data was discarded from the results. From these results could be concluded that, except for a single case, participants had sufficient visual capabilities to include their captured gaze sensor data in the analysis of gaze behavior.

4.1.2 STAI & TIS Scores

The STAI and TIS questionnaires were used to get a general idea of the participants and their fear behavior. The STAI-state was used to assess anxiety at the moment of the session taking place. It yielded a mean total score of 32.63 (σ = 8.53), and all scores were less than or equal to 50. Given that the possible scores

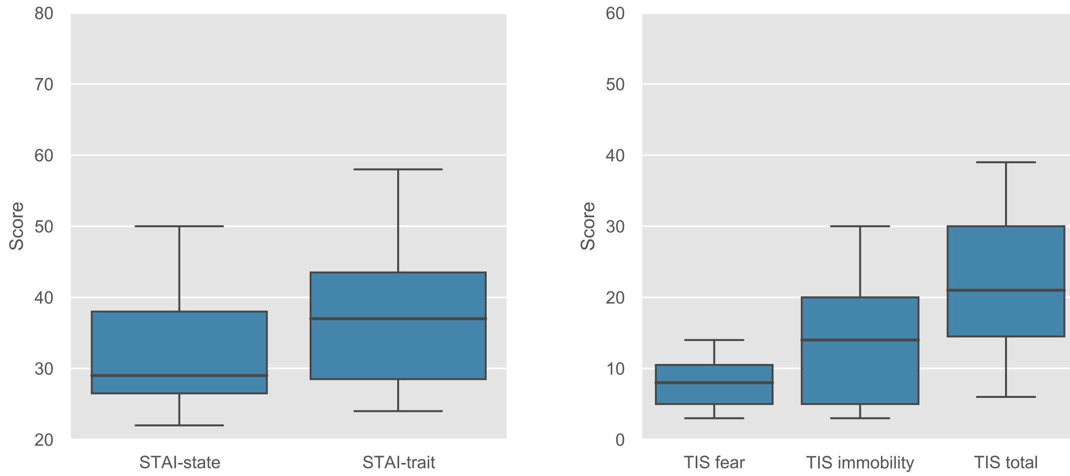


FIGURE 14: Box plots of the STAI (left) and TIS (right) scores.

lay within the range of 20-80, this means that all scores are less than or equal to the median value of this range (coincidentally 50), suggesting that these scores are relatively low. From this, it was concluded that, on average, anxiety intensity appeared to be limited before the VR session, meaning fear behavior found in the sensor data was almost certainly induced during the VR session. The STAI and TIS scores are visualized in Figure 14 and are listed per participant in Appendix A.

STAI Scores The STAI-trait was used to assess anxiety in general as a personality trait. The STAI-trait yielded a mean total score of 37.32 ($\sigma = 9.88$). Anxiety intensity in general also appeared to be limited on average, given that most scores lay below 50 too. An ANOVA on mean total STAI-state score did not yield a significant difference for gender ($F(1, 17) = .179$, $p = .677$) and an ANOVA on the mean total STAI-trait score did not yield a significant difference for gender either ($F(1, 17) = .190$, $p = .669$). Therefore, the STAI scores suggest there were no differences in anxiety between males and females.

TIS Scores The TIS was used to assess involuntary fear and immobility behavior. The TIS fear subscale yielded a mean total score of 8.11 ($\sigma = 3.60$). An ANOVA on the total TIS fear score yielded a significant difference for gender ($F(1, 17) = 5.401$, $p = .033$). The TIS immobility subscale yielded a mean score of 13.84 ($\sigma = 8.63$). An ANOVA on the total TIS immobility score did not yield a significant difference for gender ($F(1, 17) = 1.623$, $p = .220$). The total TIS yielded a mean score of 21.95 ($\sigma = 10.53$). An ANOVA on the total TIS score did yield a marginal difference for gender ($F(1, 17) = 3.437$, $p = .081$). The two TIS subscales give mixed suggestions on whether gender plays a role in involuntary fear behavior. The results show that females appear to show more involuntary fear behavior, but not for physical immobility. This difference might explain the marginal difference found for the total score, as this score is calculated by combining both subscales. This means it could be possible that more involuntary fear behavior will be shown by females in the results.

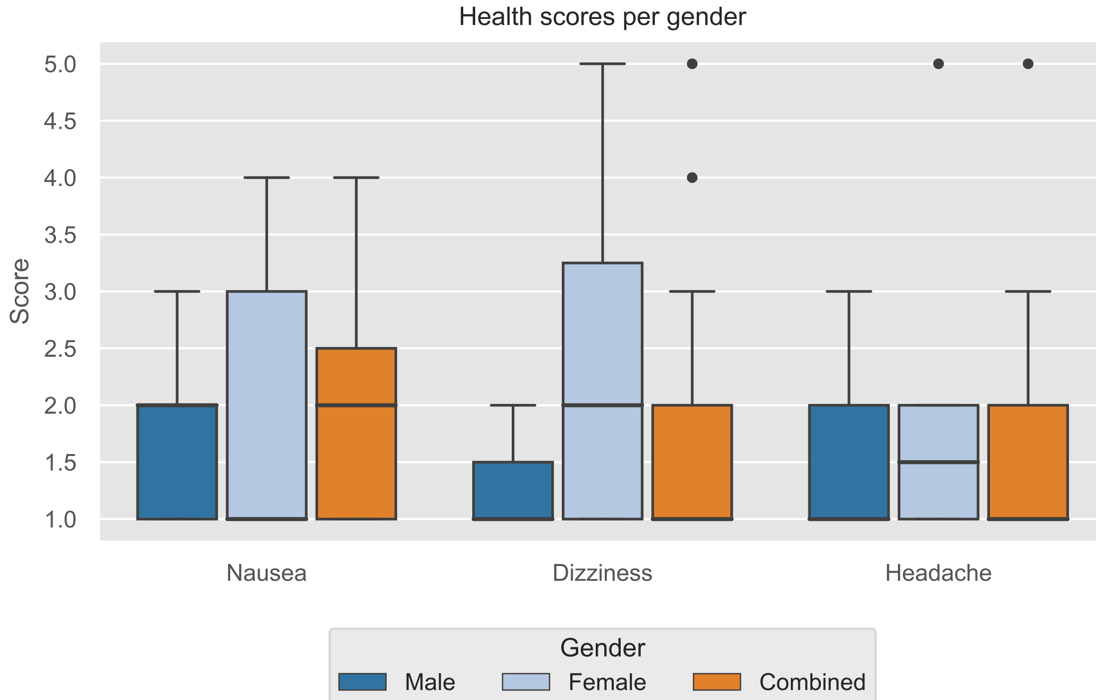


FIGURE 15: Box plots of the scores for intensities of nausea, dizziness, and headache per gender.

4.1.3 Virtual Reality Sickness Scores

Participants were asked to rate on a scale of one (not at all) to five (very much so) how sick they felt after the VR session to assess if there would be any distractions or adjusted behavior that could be caused by motion sickness, as discussed in Section 2.4.5. Participants gave a mean score of 1.84 (σ = .96) to how nauseous they felt, a mean score of 1.74 (σ = 1.15) to how dizzy they felt, and a mean score of 1.63 (σ = 1.01) to what intensity they suffered from a headache. These scores are listed in Table 7 and are visualized in Figure 15.

On average, these scores are relatively low, with their means being below a score of two. From these scores, it could be concluded that VR had a minimal negative influence on the participants' well-being. These relatively low scores may be due to the participant standing still in a single location, a minimal amount of moving elements in the VE, minimal latency of the displays, and the relatively short duration of a single session. Most of these design choices were implemented as previous works suggested taking these as precautions to reduce motion sickness, as discussed in Section 2.4.5. These design choices did appear to show their merits, as reflected by these scores.

Health Scores & Gender In Section 2.4.5 it is also discussed that a correlation exists between gender and motion sickness, where females suffer more from motion sickness compared to males. Therefore, the scores are also analyzed per gender to see if this is also the case in this research. The scores per gender are listed in Table 7 and are visualized in Figure 15 for intensity scores for nausea, dizziness, and headache.

An ANOVA on the nausea intensity scores did not yield a significant difference for gender ($F(1, 17) = .015$, $p = .903$). An ANOVA on the dizziness intensity scores did yield a significant difference for gender

		Nausea		Dizziness		Headache	
		μ	σ	μ	σ	μ	σ
Gender	Male	1.82	0.75	1.27	0.47	1.45	0.69
	Female	1.88	1.25	2.38	1.51	1.88	1.36
Combined		1.84	0.96	1.74	1.15	1.63	1.01

TABLE 7: Nausea, dizziness and headache intensity scores listed for all participants, and per gender.

		Realism of Characters		Realism of VE		Presence in VE		Immersion in VE	
		μ	σ	μ	σ	μ	σ	μ	σ
Gender	Male	2.73	1.35	2.45	1.57	2.91	1.51	2.73	1.27
	Female	3.63	1.06	3.63	0.74	4.38	0.52	4.00	0.93
VR experience	Yes	3.15	1.46	2.85	1.46	3.31	1.32	3.15	1.21
	No	3.00	0.89	3.17	1.33	4.00	1.55	3.50	1.52
Combined		3.11	1.29	2.95	1.39	3.53	1.39	3.26	1.28

TABLE 8: Realism, presence and immersion scores listed for all participants, per gender, and based on whether someone has prior VR experience.

($F(1, 17) = 5.298$, $p = .034$). When this is further inspected by looking at the average scores per gender and by looking at Figure 15, it appears that females did suffer more from nausea on average during after the VR session. An ANOVA on the headache intensity scores did not yield a significant difference for gender ($F(1, 17) = .791$, $p = .386$). These results suggest that, during this research, females did not suffer statistically differently from VR sickness compared to males based on nausea and headache intensities, except for the dizziness scores. These findings appear to be, to some extent, in line with the findings of previous works.

4.1.4 Immersion & Realism Scores

Participants were also asked to rate on a scale of one (not at all) to five (very much so) how immersed they felt and how realistic they found the VE. Participants gave a mean score of 3.53 ($\sigma = 1.39$) to what extent they felt present in the VE and gave a mean score of 3.26 ($\sigma = 1.28$) to what extent they felt immersed in the VE. Additionally, a mean score of 2.95 ($\sigma = 1.39$) was given to what extent the VE appeared realistic to the participant, and a mean score of 3.11 ($\sigma = 1.29$) was given to what extent the virtual characters appeared realistic to the participant. These scores are listed in Table 8 and are visualized in Figure 16.

From these scores, it could be determined that the participants were, on average, sufficiently positive about how immersive and realistic the VR experience was. This means that most participants were not distracted by how the VE and characters looked to a large extent, felt reasonably immersed and present in the VE, and were therefore able to show natural behavior.

Immersion Scores & Gender The immersion scores were further inspected for differences between males and females. These scores per gender can also be seen in Table 8 and are visualized in Figure 16a. An ANOVA on the realism scores of the characters did not yield a significant difference for gender ($F(1, 17) = 2.435$, $p = .137$). An ANOVA on the realism scores of the VE did yield a marginal difference for gender ($F(1, 17) = 3.771$, $p = .069$). Upon further inspection of the mean scores and by looking at Figure 16a, males gave on average lower scores to the realism of the VE than females. An ANOVA on the presence scores ($F(1, 17) = 6.827$, $p = .018$) and immersion scores ($F(1, 17) = 5.750$, $p = .028$) in the VE both yielded a significant difference for gender. Upon further inspection, again, males gave on average lower scores than females. These results suggest that, on average, females felt more immersed and present than males during the experiment. This could be motivated by comments given by some male participants after the session. For instance, a few of the male participants stated they possess a VR device or play video games frequently, which could mean they have a different or stricter perspective on the definition of a realistic VE. After all, video games are often made by companies who put a lot of effort and budget into making the environment appear as immersive and realistic as possible. Given that the environment was made by a single person and in a relatively short time, such levels of detail and realism could not be achieved for this experiment. This could be reflected in the realism scores obtained from this questionnaire for both genders combined. Other male participants stated they follow a game development-related study and that they paid attention to the appearance of various VE elements, such as small inconsistencies in texture tiling. Additionally, environments developed in Unity tend to have a specific "look" as if the objects are made of plastic, which is also slightly the case in the VE, which could reduce realism and immersion.

Immersion Scores & Prior VR Experience Lastly, a comparison for these scores was made when looking at participants who have and have not experienced VR before. These scores can also be seen in Table 8 and are visualized in Figure 16b.

An ANOVA on the realism scores of the characters did not yield a significant difference for prior VR experience ($F(1, 17) = .056$, $p = .816$). An ANOVA on the realism scores of the VE did not yield a significant effect for prior VR experience ($F(1, 17) = 0.208$, $p = .654$). An ANOVA on the immersion scores of the VE did not yield a significant effect for prior VR experience ($F(1, 17) = 0.286$, $p = .599$). An ANOVA on the presence scores of the VE did not yield a significant effect for prior VR experience ($F(1, 17) = 1.021$, $p = .327$).

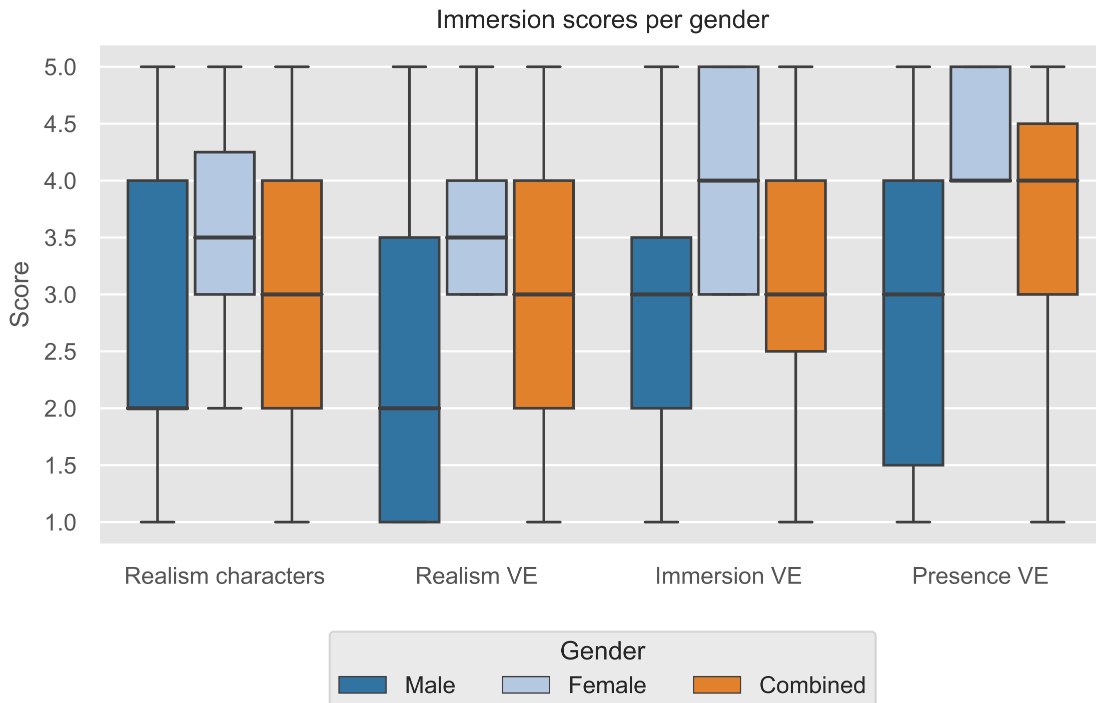
From these results could be concluded that prior VR experience did not appear to play a role in to what extent the participants felt immersed.

4.1.5 Subjective Fear Scores

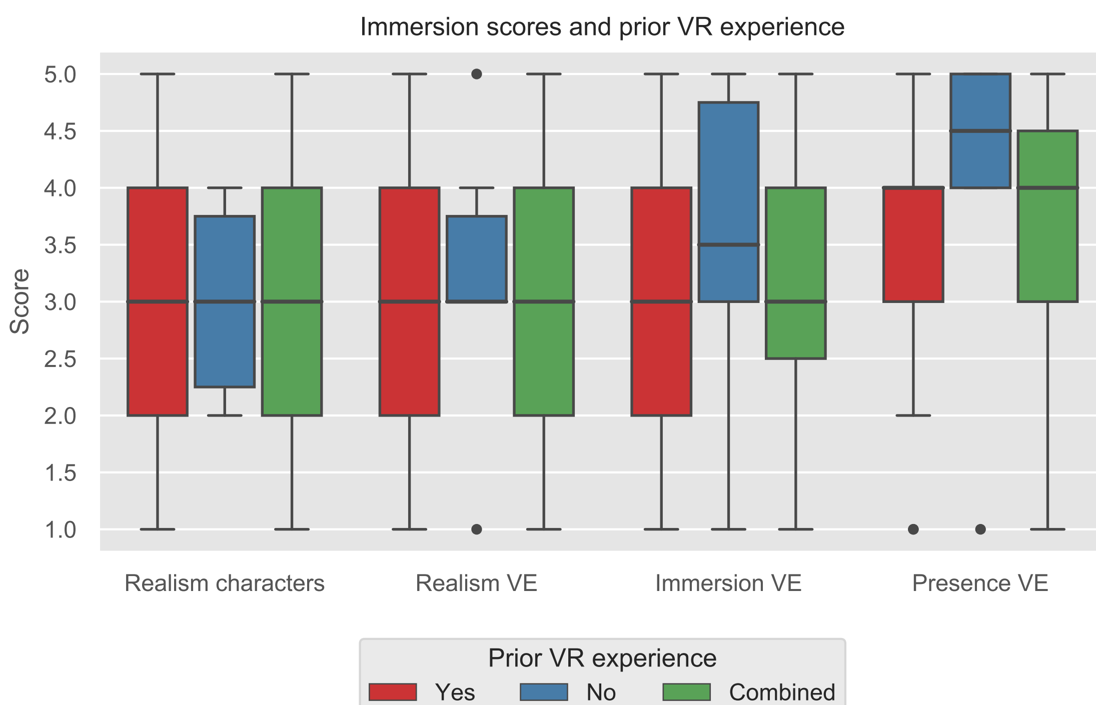
The final questions in the questionnaire were concerned with to what extent the participants were afraid or felt they froze when approached by the characters. Answers were again on a scale of one (not at all) to five (very much so). Their outcomes were used to assess if the characters' fear-inducing variants were more fear-inducing than their neutral counterparts. Additionally, they were used to assess differences in induced fear between the character models used as stimuli. These scores can be seen in Table 9.

Fear Scores per Stimulus Type To assess the differences in the intensity of induced fear, the differences between felt fear and felt freezing intensity scores were analyzed per stimulus type. These scores are visualized per stimulus type in Figure 17.

An ANOVA on the freezing intensity scores did not yield a significant effect for stimulus type ($F(1, 36) = 2.36$, $p = .133$). An ANOVA on the fear intensity scores yielded a significant effect for stimulus type



(A) Box plots of the immersion scores grouped per gender.



(B) Box plots of the immersion scores based on whether the participant has prior VR experience.

FIGURE 16: Box plots of the immersion scores of all participants combined and grouped by gender and prior VR experience.

		Freezing Neutral		Freezing Fear-Inducing		Fear Neutral		Fear Fear-Inducing	
		μ	σ	μ	σ	μ	σ	μ	σ
Gender	Male	1.00	0.00	1.64	1.03	1.00	0.00	2.27	1.27
	Female	2.25	1.16	2.75	1.49	2.13	1.13	3.25	1.49
Character model	Remy	1.44	1.01	2.50	1.43	1.11	0.33	3.00	1.63
	Josh	1.60	0.97	1.67	1.12	1.80	1.14	2.33	1.12
Combined		1.53	0.96	2.11	1.33	1.47	0.90	2.68	1.42

TABLE 9: Subjective freezing and fear intensity scores given by the participants after the VR session to the fear-inducing and neutral stimuli. Additionally, these scores can be seen grouped by gender and character model.

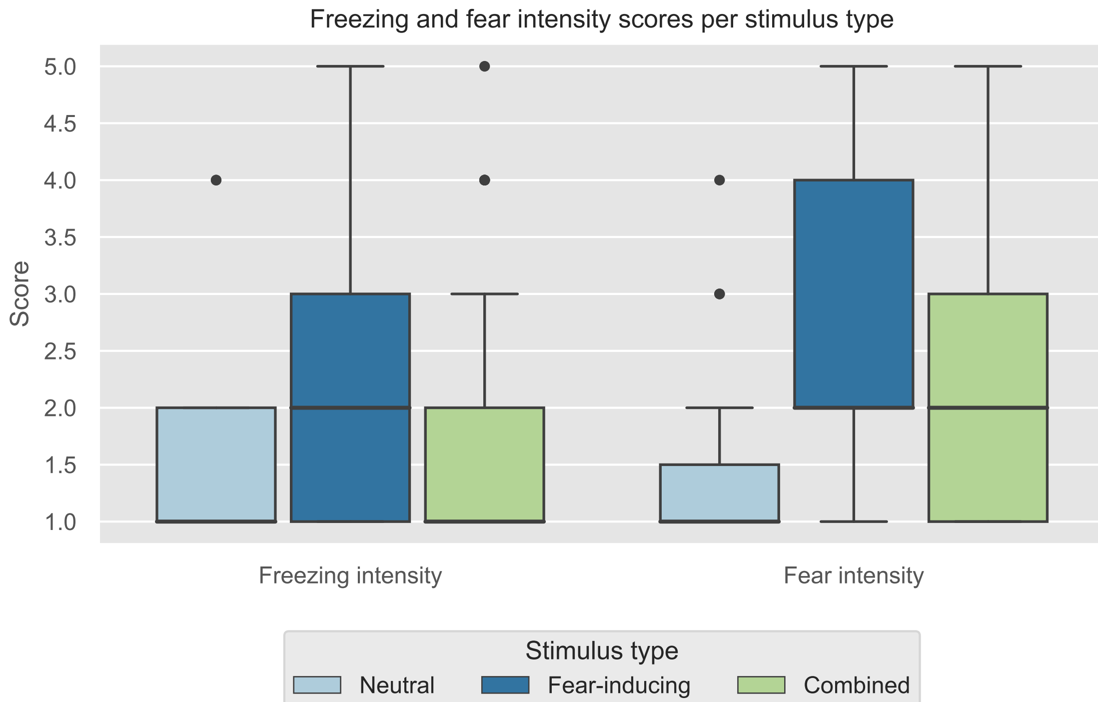


FIGURE 17: Box plots of the felt freezing and fear intensity of the participants per stimulus type.

($F(1, 36) = 9.86$, $p = .003$). Upon further inspection of the average scores and Figure 17, it appears that, on average, the fear stimuli got a higher fear-intensity score than the neutral stimuli.

Participants did not feel like they froze more when approached by the fear-inducing stimulus than the neutral stimulus. If freezing occurred (subconsciously) will be determined from the sensor data later. Participants appeared to be, on average, more afraid of the fear-inducing character variants compared to the neutral variants, which means the fear stimuli used (angry face, rifle, loud gunshot) were successfully capable of inducing fear, based on these subjective scores.

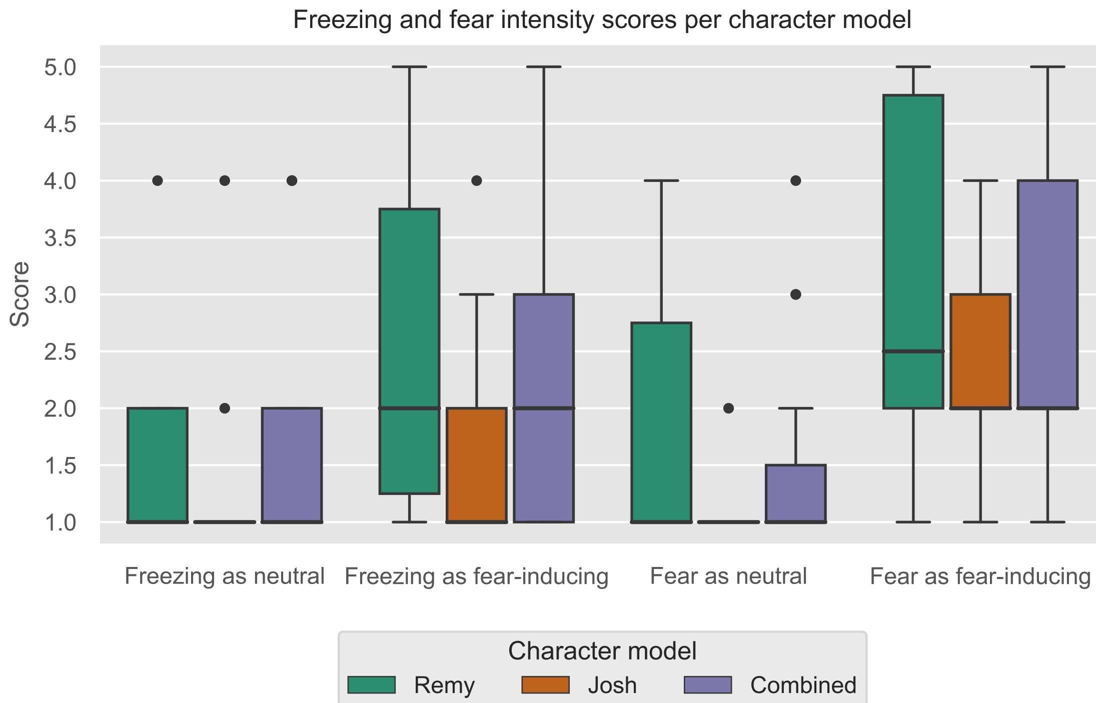


FIGURE 18: Box plots of the felt freezing and fear intensity of participants per character model. Due to lack of space, the x-labels are shortened. I.e. the left-most x-label should be read as "Freezing intensity score of this character as neutral stimulus".

Fear Scores per Character To assess to what extent both character models were able to induce identical fear or neutral behavior, the felt fear and freezing intensity scores were compared per character model. This was done to make sure the character models of Remy and Josh induce approximately the same intensities of fear as the fear-inducing and neutral stimuli. It would be ideal to have minimal differences between the two. Otherwise, these might cause different fear behavior (intensities) in the sensor data based on which character was the fear-inducing stimulus during a session. These scores are visualized per character model in Figure 18.

An ANOVA on the scores for felt freezing intensity when approached by the neutral stimulus did not yield a significant difference between the two character models ($F(1, 17) = .117$, $p = .736$). This was also the case for felt freezing intensity when approached by the fear-inducing stimulus, which also did not yield a significant difference between the two character models ($F(1, 17) = 1.962$, $p = .179$).

These comparisons were also done per character for the felt fear intensity. An ANOVA on the felt fear intensity when approached by the neutral stimulus did yield a marginal difference between the two character models ($F(1, 17) = 3.06$, $p = .098$). When this is further analyzed based on average scores and by looking at Figure 18, it appears that when Remy is the fear-inducing stimulus, the fear score for the neutral stimulus is higher on average. This means Josh is more fear-inducing as a neutral stimulus compared to Remy. Participant feedback after the sessions confirms this finding, as various participants stated they found Josh "stare right through them", which made some participants feel a little uneasy. This character model does subjectively look less friendly than Remy overall, which could also be motivated by his formal clothing and his slightly surly facial expression as his neutral variant. An ANOVA on the felt fear intensity when approached by the fear-inducing stimulus did not yield a significant difference between the two character models ($F(1, 17) = 1.053$, $p = .319$).

These scores suggest that according to the participant ratings, on average, it does not matter which character was the fear-inducing stimulus, as both appear not to show significantly different average subjective fear intensity scores. Only Josh appears to induce a bit more fear as the neutral stimulus, compared to Remy. Therefore, it is not expected to encounter significant differences in fear intensity caused by which character was used as fear-inducing stimulus during a session.

Ability to Induce Fear per Character Model To assess whether each character model induced higher intensities of fear as their fear-inducing variant compared to their neutral variant, a comparison was also made between the two variants per character model. These scores are also visualized in Figure 18.

An ANOVA on the felt freezing intensity scores yielded a marginal effect for the stimulus type of Remy ($F(1, 17) = 3.36$, $p = .084$). An ANOVA on the felt fear intensity scores yielded a significant effect for the stimulus type of Remy ($F(1, 17) = 11.54$, $p = .003$). This means that Remy appeared to induce different intensities of fear, based on its stimulus variant. Upon further inspection of the average scores and Figure 18, it appears Remy got higher fear intensity ratings as its fear-inducing variant compared to its neutral variant. The same appears to be the case for the freezing intensity ratings, though this difference is less significant.

An ANOVA on the felt freezing intensity scores did not yield a significant effect for the stimulus type of Josh ($F(1, 17) = .02$, $p = .891$). An ANOVA on the felt fear intensity scores also did not yield a significant effect for the stimulus type of Josh ($F(1, 17) = 1.06$, $p = .318$). This means that Josh did not appear to have significantly different fear and freezing intensity ratings when comparing its stimulus variants. These findings are also in line with the discussion earlier about Josh's appearance.

Fear Scores & Gender Finally, an assessment was done based on whether there are differences between the subjective fear and freezing intensity scores based on gender. These scores can be seen in Table 9 and are visualized in Figure 19.

An ANOVA on the felt freezing intensity when approached by the fear-inducing stimulus yielded a marginal effect for gender ($F(1, 17) = 3.749$, $p = .070$). An ANOVA on the felt freezing intensity when approached by the neutral stimulus yielded a significant effect for gender ($F(1, 17) = 12.950$, $p = .002$). When inspecting these further based on the average scores and by looking at Figure 19, it appears that females, on average, gave higher freezing intensity ratings compared to males for both types of stimuli. These results suggest that females felt they froze more than males, on average.

An ANOVA on the felt fear intensity when approached by the fear-inducing stimulus did not yield a significant effect for gender ($F(1, 17) = 2.374$, $p = .142$). An ANOVA on the felt fear intensity when approached by the neutral stimulus yielded a significant effect for gender ($F(1, 17) = 11.228$, $p = .004$). When inspecting this further, again by looking at the average scores and Figure 19, it appears that females, on average, gave higher fear intensity ratings when approached by the neutral stimuli. These results also suggest that males and females appeared to feel similar intensities of fear when approached by the fear-inducing stimuli.

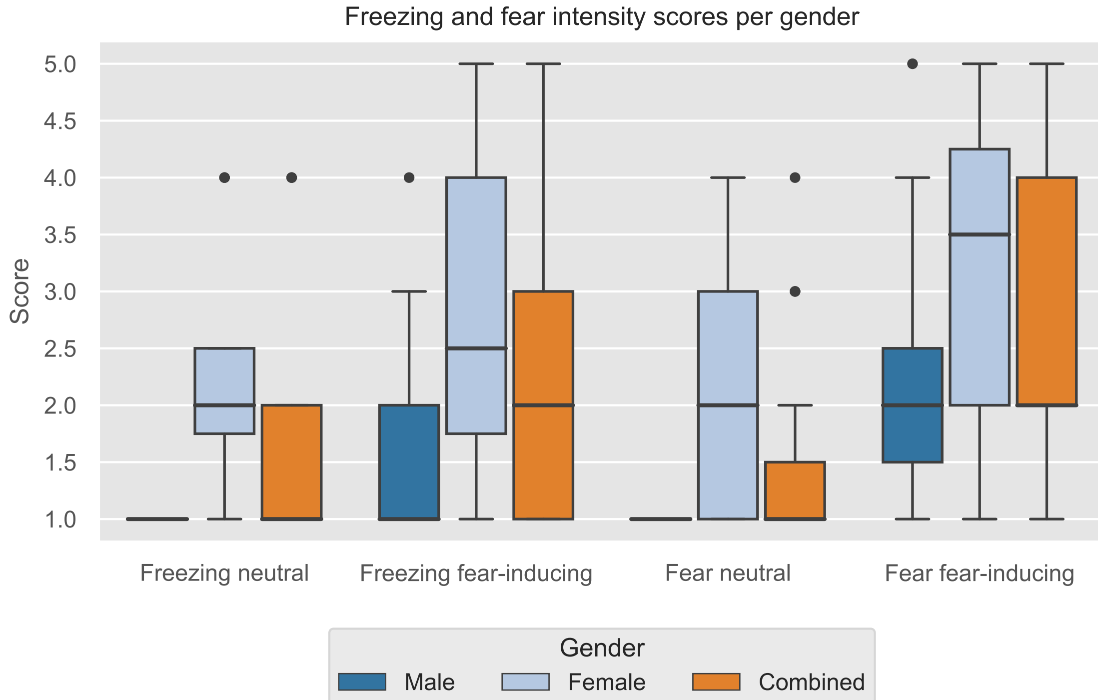


FIGURE 19: Box plots of the felt freezing and fear intensity of participants per gender.

These results suggest there are some differences between males and females when looking at the subjective fear and freezing intensity scores, where females overall appeared to have felt higher intensities of fear and felt they froze more than males when approached by the neutral stimulus. Males and females did appear to show limited differences in fear and freezing intensity when approached by the fear-inducing stimulus.

Fear Scores Between Blocks Participants were asked to rate vocally on a six-point scale of zero (not at all) to five (very much so) to what extent they were afraid of each character in-between the blocks. This was done to assess how fear manifested throughout the VR session. The average ratings per block are visualized in Figure 20.

An ANOVA on the vocal fear intensity scores yielded a significant effect for block ($F(3, 45) = 4.394$, $p = .009$). This means that fear intensity scores differed significantly per block in general. This can be seen in Figure 20 where the fear intensity scores appear to not be constant over time.

An ANOVA on the vocal fear intensity scores yielded a significant effect for stimulus type ($F(1, 15) = 21.739$, $p < .001$). Upon further inspection of the average scores, it appears that throughout all four blocks, the fear-inducing stimulus always scored higher than the neutral stimulus. This is visible in Figure 20, as the line of the fear-inducing stimulus (red) is always higher than the neutral stimulus (blue) for each block.

An ANOVA on the vocal fear intensity scores did not yield a significant effect for the interaction block $\times$ stimulus type ($F(3, 45) = .455$, $p = .715$). This can also be seen in Figure 20, as both stimulus types' fear intensity score lines show similar patterns over time (which is visualized by the black dashed line). This means that generally, fear increases and decreases similarly over time for both types of stimuli.

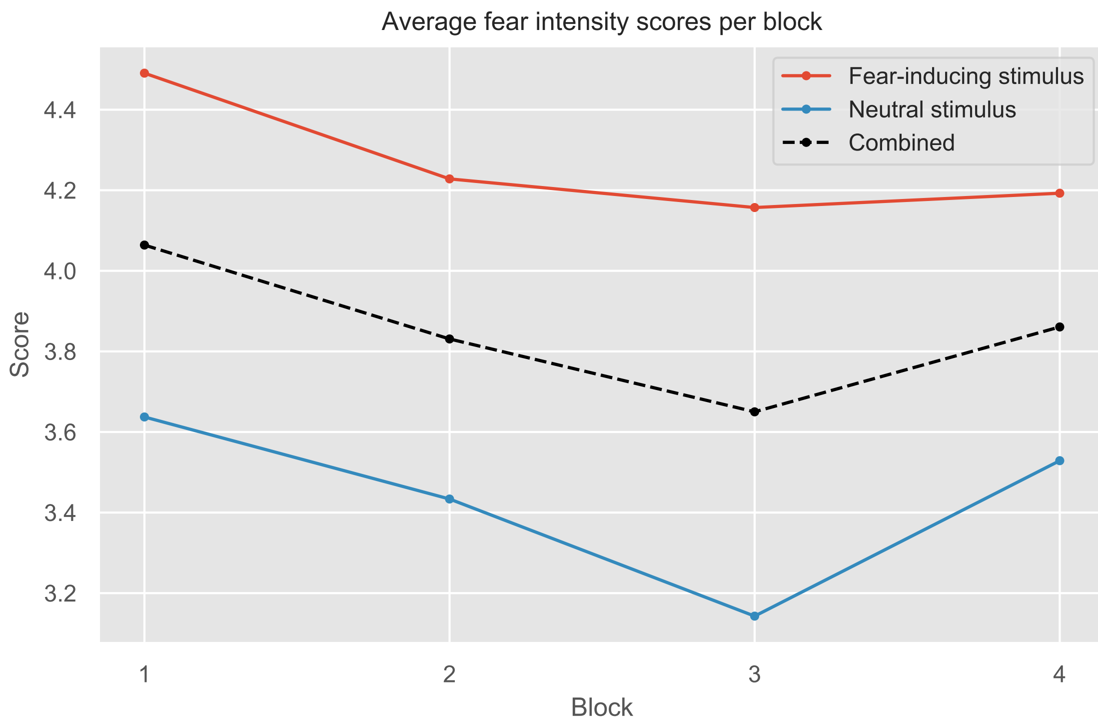


FIGURE 20: Average fear intensity scores per stimulus type, given by the participants after each block.

On average, fear scores constantly dropped during the first three blocks. During block four, fear intensity ratings appeared to increase again. Vocal feedback from participants gave some insights into why this was the case. Participants stated they got used to what was happening during the trials of the first three blocks. Therefore, the fear intensity (and thus, the scores) dropped over time. When it comes to the sudden rise during block four, participants stated they found it intimidating that both stimuli entered simultaneously and together stared at the participant. Some participants also claimed it felt like both characters were conspiring against the participant. This explains the sudden increase in fear intensity score for the neutral stimulus. The fear-inducing stimulus did not appear to become more fear-inducing in this last block, as the average growth in fear intensity score was subtle.

These subjective scores will be further used during the upcoming analyses of the sensor data to determine whether behavioral patterns could be related to felt fear intensity.

4.2 Invalid Session Data

Before proceeding to the analyses of the sensor data, some participants' sensor data had to be discarded first. In total, four sessions were deemed as partly invalid. During the first three sessions, the experiment software would not properly label the recorded sensor data, which made this data unusable. From these sessions, the filled-in questionnaire data was still usable, so only the sensor data was filtered away from the results. After these sessions took place, this error in the software was resolved. As discussed in Section 4.1.1, a fourth participant's gaze data was filtered away. Therefore, in the appendices featuring visualizations of sensor data per participant, participants P01, P02, and P03 are completely left out and P17 is only left out of the gaze data visualizations.

4.3 Body Sway

After leaving out the last interval, statistical tests were performed on the balance data measurements of the two remaining intervals to determine whether significant differences in body sway were found. Per participant, the body sway paths during block one are visualized in Appendix C. In Appendix D, the body sway measurements are also visualized per participant. In both appendices, participants were assigned a color, which makes it easier to identify individual participants and compare body sway results. The body sway analyses are only executed on block one, for the same reasons as discussed in Section 3.6.4.

4.3.1 Body Sway of Two Intervals

The following analyses were done on the two intervals of block one, which were not split up yet. The average body sway measurement values are plotted per interval and per stimulus type in Figure 21.

An ANOVA on the mean Euclidean distance did not yield a significant effect for interval ($F(1, 15) = .248$, $p = .626$), and also did not yield a significant effect for stimulus type ($F(1, 15) = 3.043$, $p = .102$). An ANOVA on the interaction interval × stimulus type did not yield a significant effect either ($F(1, 15) = .604$, $p = .449$).

An ANOVA on the mean absolute distance in ML-direction did not yield a significant effect for interval ($F(1, 15) = .233$, $p = .636$), but yielded a marginal effect for stimulus type ($F(1, 15) = 4.061$, $p = .062$). An ANOVA on the interaction interval × stimulus type did not yield a significant effect ($F(1, 15) = .312$, $p = .585$). To further analyze the marginal effect for stimulus type, it will be analyzed by looking at the average measurement values visualized in Figure 21. In this figure, it appears that trials featuring the neutral stimulus have, on average, less movement in ML-direction compared to the fear-inducing stimulus. Both lines appear to roughly be parallel to each other, suggesting a similar difference during both intervals.

An ANOVA on the mean absolute distance in AP-direction did not yield a significant effect for interval ($F(1, 15) = .300$, $p = .592$), and also did not yield a significant effect for stimulus type ($F(1, 15) = 1.894$, $p = .189$). An ANOVA on the interaction interval × stimulus type did not yield a significant effect either ($F(1, 15) = 1.066$, $p = .318$).

These results suggest that, when looking at two intervals, differences in the body sway appear to be minimal between the two stimulus types. The differences were also minimal between the intervals. Only the body sway in ML-direction appeared to be marginally larger in quantity for the fear-inducing stimulus. This means there is a possibility that participants moved more on average when approached by the fear-inducing stimulus compared to the neutral stimulus throughout the whole duration of each trial. Since there appears to be some evidence for adjusted body motion, further inspections will be done using four sub-intervals to determine if there are body sway patterns of a shorter duration.

4.3.2 Body Sway of Four Sub-Intervals

The body sway measurements will now be analyzed for the split-up intervals. The average body sway measurement values are plotted per sub-interval and per stimulus type in Figure 22.

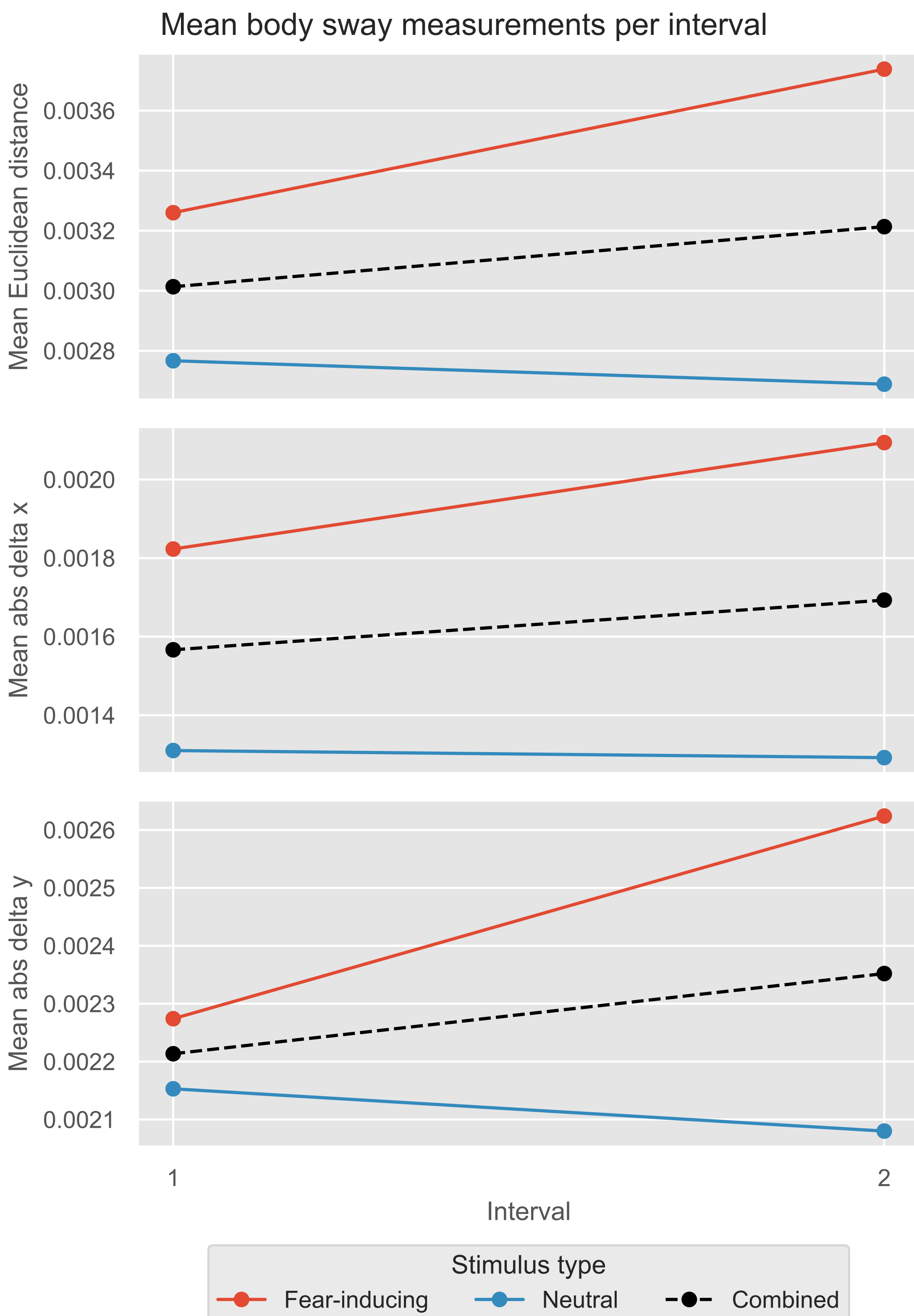


FIGURE 21: Average body sway measurement values per interval and stimulus type.

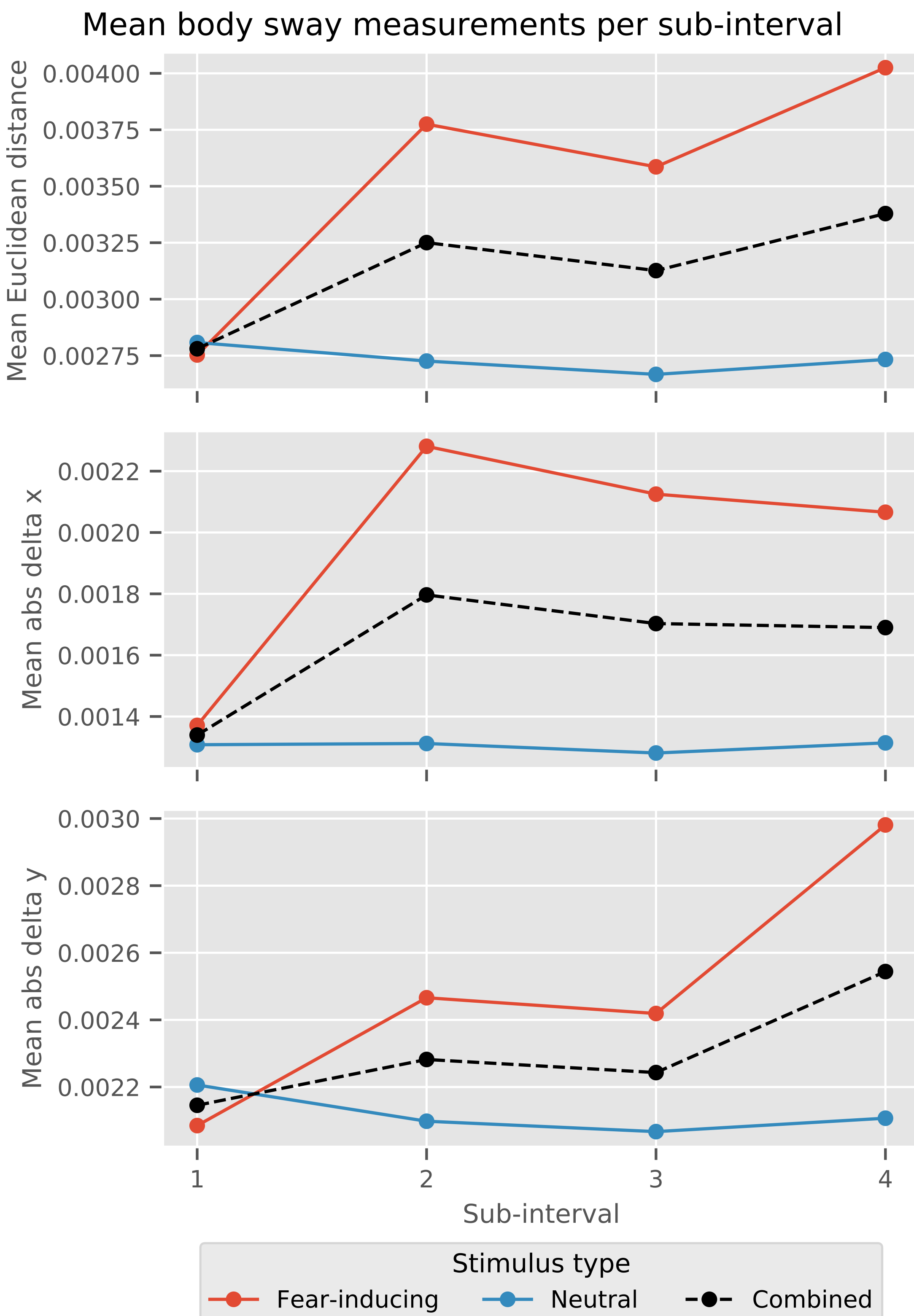


FIGURE 22: Average body sway measurement values per sub-interval and stimulus type when the data is split up into four sub-intervals.

An ANOVA on the mean Euclidean distance did not yield a significant effect for sub-interval ($F(1.511, 22.670) = 1.009$, $p = .360$). An ANOVA on the mean Euclidean distance yielded a marginal effect for stimulus ($F(1, 15) = 3.214$, $p = .093$). An ANOVA on the interaction sub-interval × stimulus type did not yield a significant effect ($F(1.713, 25.691) = 1.508$, $p = .240$). Sphericity was not assumed for sub-interval ($\chi^2(5) = 24.040, p < .001$) and for the effect sub-interval × stimulus type ($\chi^2(5) = 18.603, p = .002$), therefore the degrees of freedom were Greenhouse-Geisser corrected.

An ANOVA on the mean absolute distance in ML-direction did not yield a significant effect for sub-interval ($F(1.739, 26.085) = 1.249$, $p = .299$), but yielded a marginal effect for stimulus type ($F(1, 15) = 4.127$, $p = .060$). An ANOVA on the interaction sub-interval × stimulus type did not yield a significant effect ($F(1.896, 28.439) = .307$, $p = .307$). Sphericity was not assumed for sub-interval ($\chi^2(5) = 26.830, p < .001$) and for the effect sub-interval × stimulus type ($\chi^2(5) = 19.580, p = .002$), therefore the degrees of freedom were Greenhouse-Geisser corrected.

An ANOVA on the mean absolute distance in AP-direction did not yield a significant effect for sub-interval ($F(1.432, 21.474) = 1.117$, $p = .352$), did not yield a significant effect for stimulus type ($F(1, 15) = 2.156$, $p = .163$), and also did not yield a significant effect on the interaction sub-interval × stimulus type ($F(1.513, 22.690) = 2.135$, $p = .150$). Sphericity was not assumed for sub-interval ($\chi^2(5) = 28.136, p < .001$) and for the effect sub-interval × stimulus type ($\chi^2(5) = 21.868, p < .001$), therefore the degrees of freedom were Greenhouse-Geisser corrected.

When the marginal effect for stimulus type in ML-direction is analyzed using Figure 22, it can be seen that, on average, the fear-inducing stimulus again shows larger quantities of movement in x-direction compared to the neutral stimulus. However, this is not the case for the first sub-interval. When looking at the first sub-interval, the average quantity of movement appeared to be almost equal for both types of stimuli for all body sway measurements. During this sub-interval, the door opens, and the stimulus enters the room. Therefore, these almost equal values for both types of stimuli might be explained by the participant needing some time to perceive the stimulus entering the room.

Body Sway Excluding the First Sub-Interval From these results, it appears that during the first sub-interval, the stimulus type does not matter. Therefore, further analyses were done, leaving this first sub-interval out to determine if the remainder of the sub-intervals do show significant differences.

An ANOVA on the mean absolute distance in ML-direction did not yield a significant effect for sub-interval ($F(1.739, 26.085) = 1.249$, $p = .299$), but yielded a marginal effect for stimulus type ($F(1, 15) = 3.897$, $p = .067$). An ANOVA on the interaction sub-interval × stimulus type did not yield a significant effect ($F(1.202, 18.032) = .095$, $p = .807$). Sphericity was not assumed for sub-interval ($\chi^2(2) = 13.442$, $p = .001$) and for the effect sub-interval × stimulus type ($\chi^2(2) = 15.257, p < .001$), therefore the degrees of freedom were Greenhouse-Geisser corrected.

An ANOVA on the mean absolute distance in AP-direction did not yield a significant effect for sub-interval ($F(2, 30) = 1.134$, $p = .335$) and also did not yield a significant effect for stimulus type ($F(1, 15) = 2.724$, $p = .120$). An ANOVA on the interaction sub-interval × stimulus type did not yield a significant effect either ($F(1.341, 20.117) = 1.135$, $p = .319$). Sphericity was not assumed for the effect sub-interval × stimulus type ($\chi^2(2) = 9.461$, $p = .009$), therefore the degrees of freedom were Greenhouse-Geisser correct.

Since both measurements show similarities with statistical tests performed on all four sub-intervals, further statistical tests on the Euclidean distance were omitted. These results suggest that both sub-interval and stimulus type appear to have no significant influence on any of the body sway measurements, except for stimulus type in ML-direction, where the fear-inducing stimulus caused a larger quantity of movement in this direction compared to the neutral stimulus. This appears to be the case when both including and leaving out the first sub-interval.

4.3.3 Center of Balance Patterns

Figure 35 visualizes the 2D body sway (or COB paths) per participant and trial type from a top-down perspective of the balance board. These plots give a general overview of body sway behavior per participant. These plots could provide insights into how participants generally behaved and could showcase some individual differences between participants, which could not be extracted from the statistical tests.

Participants appeared to all show different quantities of body movement. When comparing the neutral and fear-inducing trial paths per participant, participants were grouped based on whether they moved more or less during the fear-inducing trials compared to the neutral trials. This was based on the amount of spread of the COB paths. Occasional loops or outliers were ignored in this visual inspection, so this grouping was mostly based on the overall shape of the paths and was subjectively done.

Increased Body Sway During Fear-Inducing Trials The first group consisted of participants who generally appeared to move more during the fear-inducing trials than during the neutral trials, which are participants P04, P05, P06, P08, P09, P11, P13, P16, and P17. These participants appeared to show increased body motion, which resembles the correlates of anxiety behavior more than fear or freezing behavior, as discussed in Section 2.3.2. Vocal fear intensity scores of block one (Table 11) suggest that most participants of this group gave higher fear-intensity scores to the fear stimulus than the neutral stimulus. P06 and P11 were, to the same extent, afraid of the fear-inducing and neutral stimuli, but P06 gave lower scores than P11. It appears that P06 moved only subtly more during the fear-inducing stimuli, while P11 moved visibly more, which might suggest that induced fear behavior from the fear-inducing stimulus appears to be subconscious, given these participants rated both stimulus types similarly. P08 appeared to barely feel afraid, given the vocal fear-intensity scores of one and zero for the fear-inducing and neutral stimulus, respectively. However, a subtle difference in path spread can be found for this participant, ignoring the loop in the paths of the fear-inducing trials. When the COB paths are related to the fear intensity scores, this group's participants suggest that, in general, participants move more when their fear intensity was higher.

Reduced Body Sway During Fear-Inducing Trials The second group consisted of participants who generally appeared to move less during the fear-inducing trials than during the neutral trials, which are participants P07, P10, P12, P14, P15, P18, and P19. This group of participants appeared to show reduced body motion when approached by a fear-inducing stimulus, which is in line with the found correlates of freezing behavior in humans from related literature, as discussed in Section 2.2.1. P10, P14, and P18 gave zeroes and ones to both types of stimuli, suggesting they claimed not to experience fear. However, they did appear to show more condensed movement during the fear-inducing trials, which might suggest that freezing behavior is induced subconsciously in these participants. P15 appeared to be the only participant who gave a higher score to the neutral stimulus than the fear-inducing stimulus (during block one), so this participant again shows evidence of increased body sway when the given fear intensity score is higher. P19 is an example who gave relatively high fear intensity scores to both types of stimuli, while the score was higher for the fear-inducing stimulus (five versus three). This participant's body sway appeared to be more condensed in ML-direction. When the COB paths are related to the fear intensity scores, his group of participants suggests that, in general, participants moved less when their fear intensity was higher.

4.3.4 Body Sway Patterns Over Time

The body sway measurements did not appear to show large or statistically significant differences per sub-interval or stimulus type. Additionally, the general COB patterns showed conflicting patterns in body sway, where some participants moved more during the fear-inducing trials and the other participants moved less during these trials. Therefore, it could be beneficial to analyze the body sway measurements over time to look for deviating behavior of shorter duration or reduced intensity. When looking at the body sway measurements plotted in Figure 36, there are some overall differences in body sway amount between the types of trials. Generally, the body sway measurements appear to be more evenly spread during the neutral trials, especially in ML-direction. These plots contain much fewer and less prominent peaks than the fear-inducing trials, which might suggest that, generally, more body sway occurred during the fear-inducing trials. Various prominent peaks in the signals can be found for both types of trials. These peaks could suggest some adjusted behavioral patterns but could also be caused by coincidental body motions or certain trial events. These patterns will now be discussed in more detail.

Ignoring Trial Start Any peaks in the range of 0-5 seconds will either be seen as coincidental movements or will be ignored in the following analyses. None of the stimuli were presented during these times, and it is unexpected that fear behavior induced during a trial will occur at the start of the subsequent trial, as fear behavior is of relatively short duration, as discussed in Section 2.3.2.

Participants With Large Quantities of Movement An initial global inspection shows that participants P05, P11, P15, P17, and P18 show relatively large quantities of movement in the plots compared to the other participants. This can be seen by their large number of peaks and by their relatively large sizes. When looking at their general COB path shapes in Figure 35, they appeared to overall move relatively more during both types of trials than other participants. Hence these participants' signals tend to overshadow other participants in these plots.

When looking at the plots of the neutral trials' body sway measurements, P15 appears to show relatively large quantities of movement, especially around the 11-12 seconds range and in AP-direction, compared to the other participants. Some peaks of this participant are also visible during the fear-inducing trials, especially around the trial end after roughly 17 seconds. P15 appears to be slightly more afraid of the neutral stimulus, as was found earlier from the vocal fear intensity scores. However, this difference is minimal, and the scores are generally on the lower side. STAI and TIS scores also did not suggest this person was subject to high intensities of fear or was a fearful person in general (Figure 10). During the session, it was noticed that the participant wavered on the balance board. Therefore, the relatively large number of peaks throughout the neutral trials could be caused by natural movement behavior or coincidental movements of this participant, which were not motivated by fear.

Body Sway at Door Opening Event When looking at the plots of both types of trials, an interesting difference is visible at the door opening event, which takes place after five seconds. Shortly after, various large quantities of movement can be seen for the fear-inducing trials in the range of 5-12 seconds, which is almost not the case for the neutral trials, where much more occasional and subtler peaks occur during this period. Participants who showed the highest amount of movement during the fear-inducing trials in this period were participants P05, P11, P13, and P17. P09 additionally appeared to only show peaks around six seconds in this range. The peaks at the range of 5-8 seconds could also be seen, to a lesser extent, in the plots of the neutral trials. These peaks could be motivated by the participant needing some time to perceive the stimulus or turn the body and head towards the stimulus.

As stated earlier, P11 was afraid of both characters to the same extent. Therefore, it is expected that this participant showed similar quantities of movement during both trial types, which eventually was not the case for this range of time. This might suggest the peaks of this participant were coincidental movements. P05's peaks might also be caused by coincidental movements, which might be explained by the loops in positive ML-direction in the fear-inducing trial COB paths. P17 already appeared to move a lot in general, but the sudden large quantities of movements at the door opening event are remarkable. This might be explained by the participant needing to move more to adjust to perceive the stimulus, but a clear reason could not be found. Given that the stimulus is far away or not properly visible yet at the door opening event, these behavior patterns instead suggest they were not caused by fear. Given that P05, P11, P13 appear to have some outliers or loops in their COB paths, most peaks might also be explained by coincidental movements.

Body Sway at Gun Shot Event At the range of 13-16 seconds, which is the period in which the fear-inducing stimulus fires the rifle, the body sway measurements appear to show larger quantities of movement for the fear-inducing trials, as there are peaks visible for multiple participants. This could be explained by participants getting scared by the loud rifle shot sound effect being played through the headphones. During these events, when observing the participants, it was noticeable that most participants were startled when the shot took place. This startle response was stronger during the earlier fear-inducing trials, especially at the first time the shot took place, hence the larger number of peaks during this period for the fear-inducing trials.

Body Sway at Trial End For both types of trials, the body sway measurements appear to show a sudden rise in the number of peaks during the last few seconds of the signals, which are the periods in which the screen is gray, and the participant is not able to see anything. The number of peaks appears to be lower for the fear-inducing trials, and these peaks appear to be less prominent compared to the neutral trials. This might suggest that, generally, freezing behavior could occur during the gray area after the shot had taken place. However, it is uncertain to what extent the screen being gray influenced body sway behavior. For instance, participants could be slightly disoriented, so they had to adjust their movement behavior not to lose balance, which could coincidentally have happened more often during the neutral trials. Therefore, there is not much strong evidence that could explain the differences in body motion at the trial end.

4.3.5 Discussion of Body Sway

The statistical tests from Section 4.3.1 and Section 4.3.2 overall suggest that no major differences in average body sway exist between the trials of the two types of stimuli. A marginal effect was found that participants moved more on average in ML-direction when a fear-inducing stimulus was presented, also when the first sub-interval was excluded. The COB paths in Section 4.3.3 suggested that body sway behavior was different between participants. When comparing the COB paths for the fear-inducing and neutral trials, it became clear that the participants followed two general types of behavior: participants either moved more or less during the fear-inducing trials compared to the neutral trials. For a large portion of participants, the vocal fear intensity scores appeared to be related to what extent their body sway behavior was adjusted during the fear-inducing trials. Other participants who claimed not to be afraid did appear to show adjusted behavior during the fear-inducing trials, which suggests this adjusted behavior was induced subconsciously. Since some of those participants appeared to move less during the fear-inducing trials, these participants showed most evidence of freezing behavior. Section 4.3.4 looked further into body sway over time to determine whether more subtle or short-lived types of behavior occurred during the trials. Overall, it appeared that movement intensities appeared to be reduced and more spread out during the neutral trials, where the fear-inducing trials had more sudden and larger quantities of movements at times. Larger quantities of movement appeared during the door opening event for the fear-inducing trials. However, these were mostly motivated by coincidental movements. Larger quantities of movement also appeared when the gunshot took place during the fear-inducing trials, as the gunshot made many participants startle to some extent. Body sway appeared to be reduced during the trial end of the fear-inducing trials, which might suggest that freezing behavior occurred here, though the evidence of this is not strong.

The results from these analyses combined do not give a general and clear answer whether body sway was different between the types of stimuli presented. Generally, it appears more movement occurred during the fear-inducing trials, which is more a characteristic of anxiety behavior than fear behavior, as discussed by related works in Section 2.3.2. However, when looking at individual participants, there were multiple cases of participants who showed reduced body sway behavior, which appear to show more characteristics of freezing behavior, as discussed by related works in Section 2.2.1 and Section 2.2.4. When considering the measurements over time, it could also be possible that the coincidental movement peaks or the startle effect of the rifle shot might have influenced the results of the statistical analyses to a large extent, which might have made it more difficult to find evidence of freezing behavior in these results. To what extent this is the case, is uncertain.

4.4 Gaze Behavior Single Stimulus

After analyzing body sway, further analyses were done on the gathered gaze data. The gaze data consisted of gaze targets and pupil sizes over time. These will be analyzed to determine if differences in gaze behavior were found based on the type of stimuli presented.

4.4.1 Gaze Targets of Block One

First, the gaze targets of block one were analyzed, as the participants were required only to look straight ahead to perceive the stimuli, which did not require too many adjustments of the body and gaze behavior to perceive them, resulting in the most clean gaze data to use for the analyses.

Grouping Environmental Gaze Targets Certain gaze targets in the following analyses were grouped, as these were less important than others. For instance, all environmental elements such as doors, walls, the ceiling, and the floor were grouped as a single *Environment* gaze target. It would not be required to know at which door or wall the participant was looking since these were not part of the fear-inducing or neutral stimuli.

Important Gaze Targets From an initial inspection of the gaze target data, it became apparent that certain body parts were more looked at than others, such as the head. The torso appeared to be a common gaze target too. However, little to no attention was paid to the limbs. The weapon also appeared to be a common gaze target during the fear-inducing trials. Therefore, it was decided that the head and weapon were important gaze targets to analyze. This is also motivated by that the face and weapon were used to induce fear. In Appendix E.1, the gaze targets can be seen per participant and per trial type over time for block one.

Head The head (or face) appeared to be an important element of the stimulus, as it was often looked at. The face could have either an angry or neutral expression. Therefore, it was interesting to analyze if one of the facial expressions drew more attention. The following statistical tests will focus on the time spent looking at the *Head* gaze target. In these statistical tests, the original two intervals will be used to get a general idea of the gaze behavior over time. The average time spent looking at the head per interval and stimulus type is visualized in Figure 23.

An ANOVA on the mean time spent looking at the head yielded a significant effect for interval ($F(1, 14) = 26.582$, $p < .001$), and also yielded a significant effect for stimulus type ($F(1, 14) = 10.767$, $p = .005$). An ANOVA on the interaction interval × stimulus type did not yield a significant effect ($F(1, 14) = .170$, $p = .687$). When looking at the mean times spent looking at the head per interval, it can be seen that the time spent looking at the head is, on average, longer during the second interval, which makes sense as the stimulus is much closer to the participant and takes up more space within the FOV. This effect is also visible when the time spent looking at the head is plotted per stimulus type. Here the time spent looking at the head of the neutral stimulus is longer compared to the fear-inducing stimulus.

It appears that gaze is more directed towards the head of the neutral stimulus than the fear-inducing stimulus. To determine if the gaze is directed towards a different fear-inducing gaze target, the same statistical tests were done, also including the time spent looking at the weapon.

Head & Weapon The time spent looking at the weapon and head will be combined for the following statistical tests. The time spent looking at the rifle during neutral trials is always zero, as the stimulus does not hold a weapon. Therefore, for the neutral stimulus, the combined time is always equal to the time spent looking at the head. The average time spent looking at the head and weapon combined is visualized per interval and stimulus type in Figure 24.

An ANOVA on the mean time spent looking at the head and weapon combined yielded a significant effect for interval ($F(1, 14) = 67.709$, $p < .001$), and also yielded a significant effect for stimulus type ($F(1, 14) = 14.396$, $p = .002$). An ANOVA on the interaction interval × stimulus type also yielded a

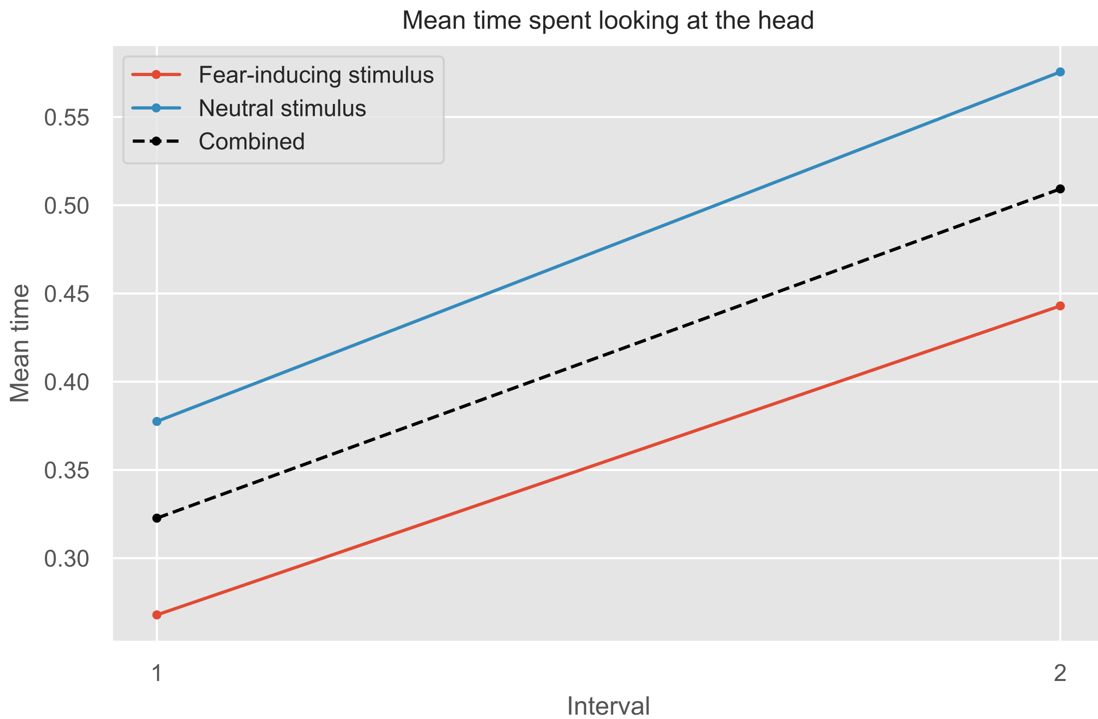


FIGURE 23: Average time spent looking at the head gaze target per interval and stimulus type.

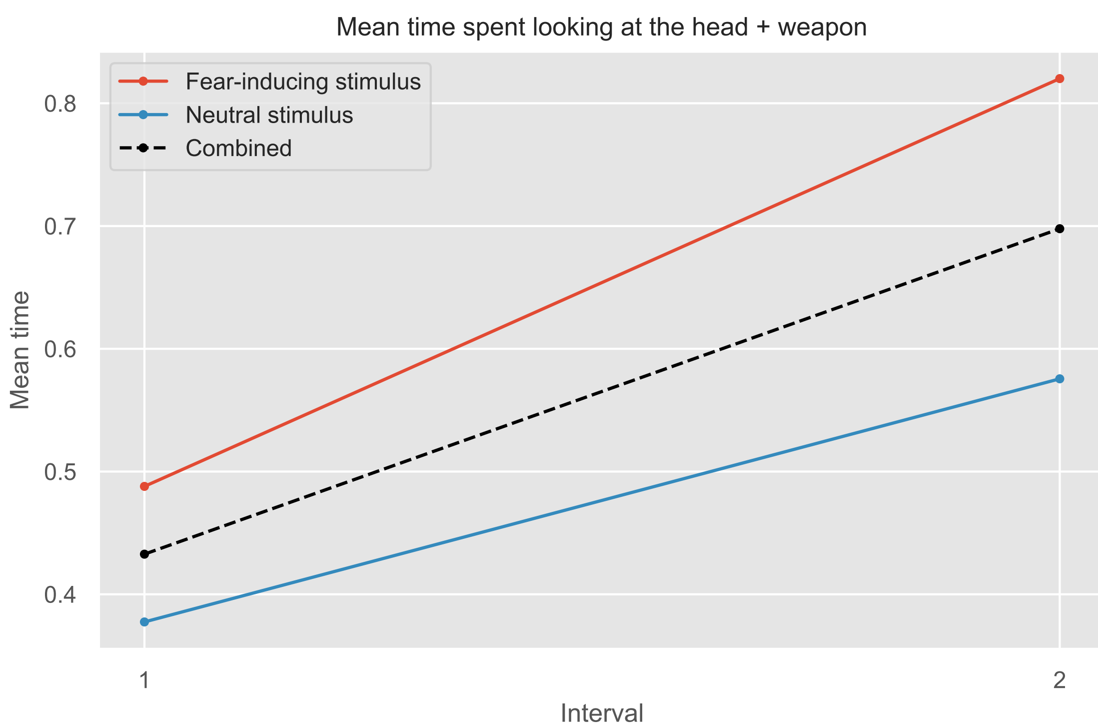


FIGURE 24: Average time spent looking at the head + weapon gaze targets per interval and stimulus type.

significant effect ($F(1, 14) = 5.802$, $p = .030$). Again, it can be seen that the time spent looking at the head and weapon is longer for the second interval, for the same reason as before. For stimulus type, the results are now flipped, where the average time spent looking at the head and weapon of the fear-inducing stimulus is now larger than for the neutral stimulus. When analyzing the interaction effect, it appears that the time spent looking at the fear-inducing stimuli increases relatively stronger in the second interval compared to the neutral stimuli. This could be motivated by the weapon's gaze target taking up a lot of space in the FOV of the participant, especially when the stimulus stands close by the participant. The gaze targets of the fear-inducing stimulus combined cover up more space in the FOV than the neutral stimulus.

As it appears from the results of the analyses of the head with and without the weapon, gaze appears to be more directed towards the fear-inducing stimulus when the weapon is included, suggesting that gaze tends to direct towards the weapon.

To determine whether the left out body parts played a role in gaze behavior, some further statistical tests were done on the entire character's body, including the torso and limbs. Due to the weapon partly overlapping the body of the fear-inducing stimulus from the participant's point-of-view, this might lead to incorrect results suggesting participants looked more often at the neutral stimulus. Therefore, the decision was made to perform these tests on the character and weapon gaze targets combined, rather than just the stimulus character.

Character & Weapon The following statistical tests were done, including the full character body and the weapon. The average time spent looking at the character and weapon combined is visualized per interval and stimulus type in Figure 25.

An ANOVA on the mean time spent looking at the full character and weapon combined yielded a significant effect for interval ($F(1, 14) = 159.153$, $p < .001$). The effects were nonsignificant for stimulus type ($F(1, 14) = 2.035$, $p = .176$). An ANOVA on the interaction interval × stimulus type did not yield a significant effect ($F(1, 14) = .000$, $p = .997$). Yet again, it can be seen that the time spent looking at the combined stimulus type is longer for the second interval, for the same reason as before. From these results, it can be seen that there is no statistical difference between the average time spent looking at the different stimulus types when including all body parts and the weapon.

From these final statistical tests, it appears that the average time spent looking at both types of stimuli is identical. However, the other tests only looking at the head with and without the weapon suggest the distribution of our attention appears to be different, where participants appeared to look more at the face of the neutral stimulus and more at the weapon of the fear-inducing stimulus. To determine if any temporal gaze patterns emerge, some further visual inspections of the gaze data will be done on the split-up data.

Gaze Patterns Over Time In Figure 37, the gaze targets per interval and participant are visualized for block one. Note that in this figure, the two intervals are split up into four intervals. Overall it appears that less time was spent on looking at the environment during the fear-inducing trials compared to the neutral trials. At the start of the trials, the environment is more looked at compared to the other intervals of the trials. This could have either been caused by the gaze tracker being less accurate when the stimulus was farther away or by the participant still looking around and needing some time to adjust to perceive the stimulus. The participants also appear to look more often at the body of the neutral stimulus compared to the fear-inducing stimulus.

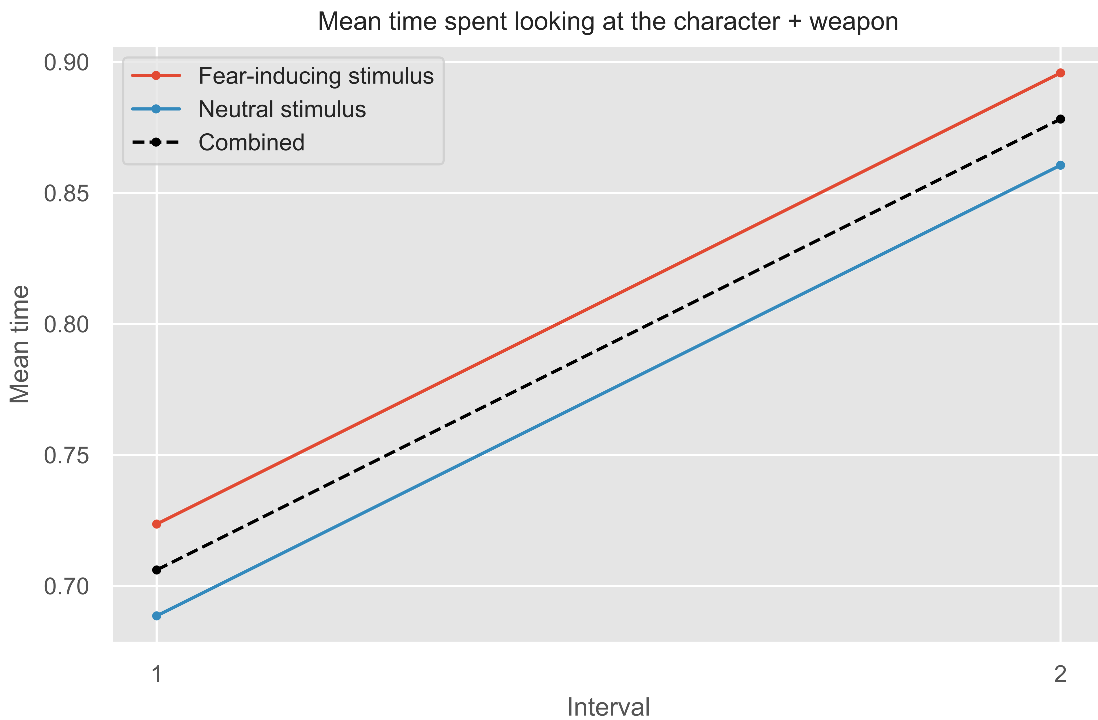


FIGURE 25: Average time spent looking at the character + weapon gaze targets per interval and stimulus type.

Here it can also be seen that less time is spent looking at the head during the fear-inducing trials compared to the neutral trials. Instead, mostly, this time is spent looking at the weapon.

These patterns could suggest that participants look around more and centralize their gaze less to the neutral stimuli compared to the fear-inducing stimuli. During the neutral trials, the gaze appears to explore more parts of the VE and other body parts than just the head. These findings are in line with the statistical tests.

Most participants showed these general patterns to some extent. However, some differences are visible in the proportions of the gaze targets between participants. For instance, some participants appeared to almost entirely fixate on the head, almost completely ignoring the environment or body, during the neutral trials (such as P04). During some of the intervals, others appeared to look more often at the body than the head (such as P10 and P11). During the fear-inducing trials, some participants looked more at the head than the weapon (such as P05, P06, and P07), and some hardly looked at the head (such as P14). P19 stood out the most, as this participant appeared to show quite exploratory behavior during the fear-inducing trials, unlike other participants. It is unclear why this participant appeared to show such deviating patterns. A possible explanation is that the participant was trying to evade the threat by looking away. The relatively large amount of time the gaze target was *Nothing* (being the largest category in the last interval) might also suggest the gaze tracker was malfunctioning or lost track of the pupils relatively often. This participant gave the fear-inducing stimulus the highest possible fear rating (five), which was the only case of such a high rating, so perhaps this high intensity of fear might also have caused this behavior.

As expected, the time spent looking at the stimuli overall increases over time and the time spent looking

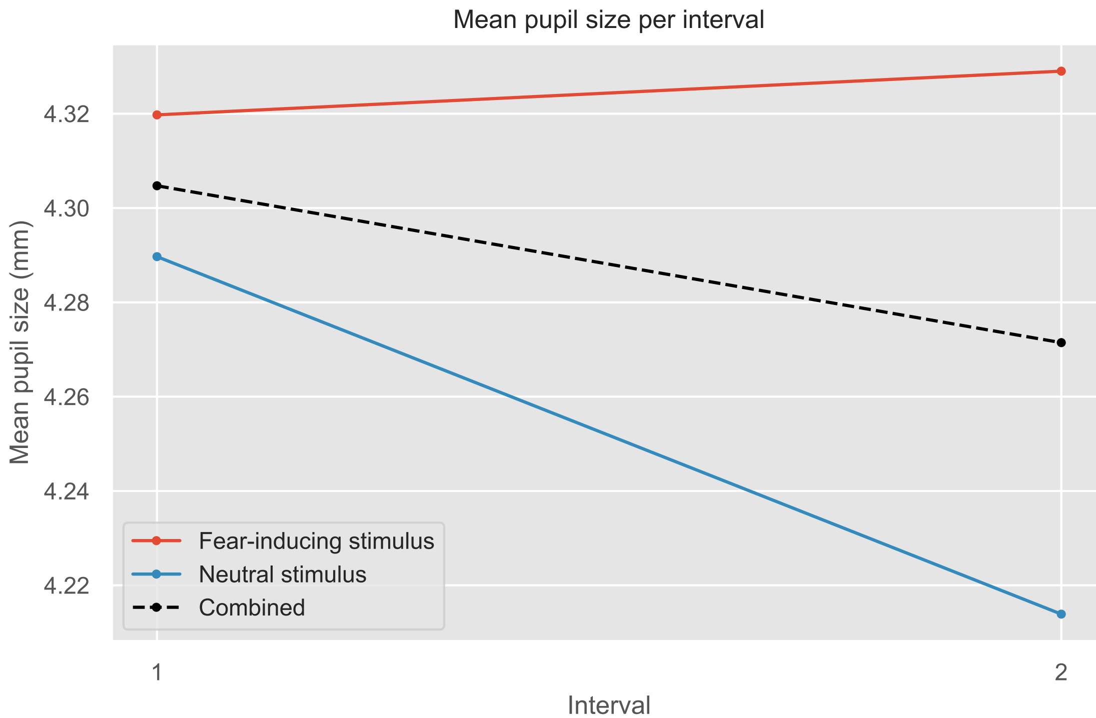


FIGURE 26: Average pupil size (mm) per interval and stimulus type.

at the environment reduces, which was explained by the stimuli taking up more space in the FOV of the participant the nearer they are located to the participant.

Participants P05, P07, P10, P14, and P18 gave relatively low fear intensity scores between the blocks (two or lower). However, their behavior did not seem to deviate too much from the general patterns discussed earlier. For instance, they all still appear to look often at the weapon. Therefore, it appears that this behavior is subconscious.

4.4.2 Pupil Size

Now that gaze targets have been analyzed, we proceed with analyses on the pupil size. When statistically testing the pupil size, the eye (left, right) has been included as a between-subjects factor, besides interval and stimulus type. Again, these tests were done at two intervals. Average pupil sizes per interval, stimulus type and eye are visualized in Figure 26 and Figure 27.

An ANOVA on the mean pupil size did not yield a significant effect for interval ($F(1, 14) = .893$, $p = .361$), did not yield a significant effect for stimulus type ($F(1, 14) = 2.069$, $p = .172$), did not yield a significant effect for eye ($F(1, 14) = .244$, $p = .626$), and also did not yield a significant effect for the interaction interval × stimulus ($F(1, 14) = 2.429$, $p = .141$). The interaction interval × eye yielded a marginal effect ($F(1, 14) = 3.911$, $p = .068$). The interaction stimulus type × eye did not yield a significant effect ($F(1, 14) = .001$, $p = .974$). Finally, the interaction interval × stimulus type × eye did not yield a significant effect ($F(1, 14) = 1.326$, $p = .269$).

From these results, it can be concluded that pupil size does not appear to change significantly depending on

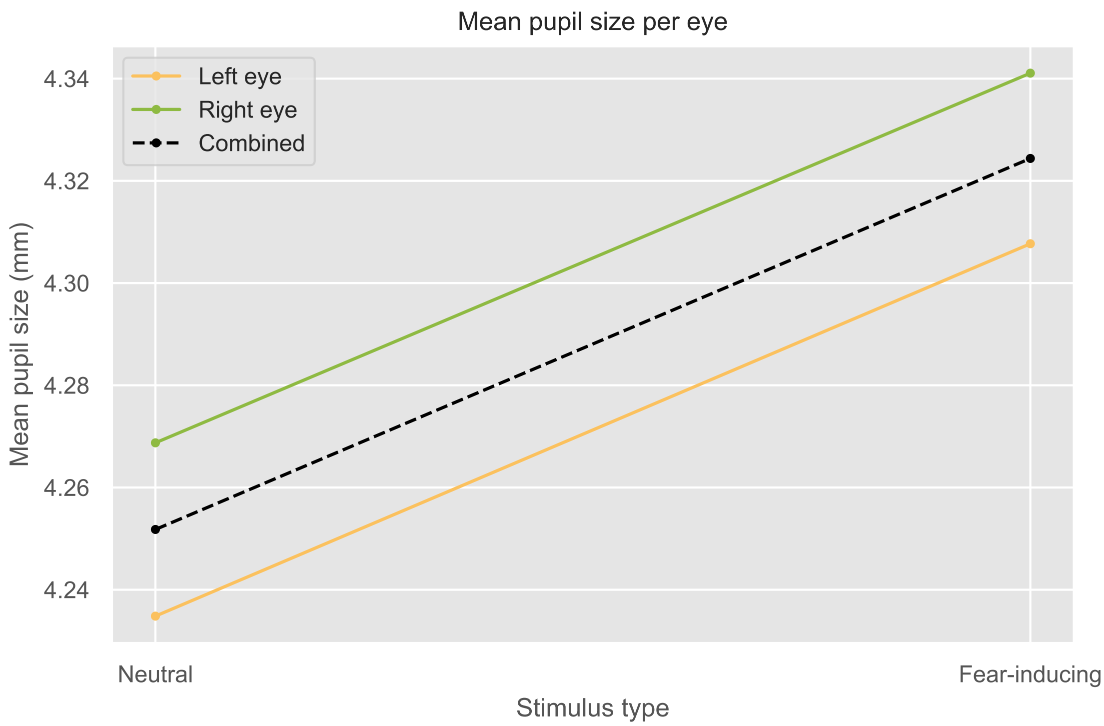


FIGURE 27: Average pupil size (mm) per eye and stimulus type.

the interval or the type of stimulus. The marginal effect found for the interaction interval × eye is challenging to explain, as it is also not expected that one of the eyes would behave significantly different compared to the other. Figure 27 suggests that the differences in pupil size between the two eyes were minimal. Therefore this marginal effect appears to not be of high interest.

Since the results of pupil size do not appear to be significantly different based on the stimulus type and interval, the pupil size will not be further analyzed for the other blocks.

4.4.3 Gaze Targets of Multiple Blocks

The gaze targets were analyzed for block one, meaning gaze behavior has been analyzed when the stimulus could only approach from straight ahead of the participant. To determine whether gaze behavior is adjusted when the stimuli approach from different angles, further statistical analyses were performed on the average time spent looking at the full character and weapon combined, including blocks 2 and 3. Block (1, 2, 3) will now be included as a within-subjects factor for the following statistical tests, and the trials are split up again into four sub-intervals to have a more detailed overview of gaze behavior over time. The average time spent looking at the character and weapon combined are visualized per block in Figure 28 and per sub-interval of the three blocks combined in Figure 29. The gaze targets of blocks two and three are also visualized per participant in Appendix E.2 and Appendix E.3, respectively.

An ANOVA on the mean time spent looking at the full character and weapon combined yielded a significant effect for block ($F(2, 28) = 35.767$, $p < .001$), and for sub-interval ($F(3, 42) = 131.982$, $p < .001$). An ANOVA on the mean time spent looking at the full character and weapon combined did not yield a significant

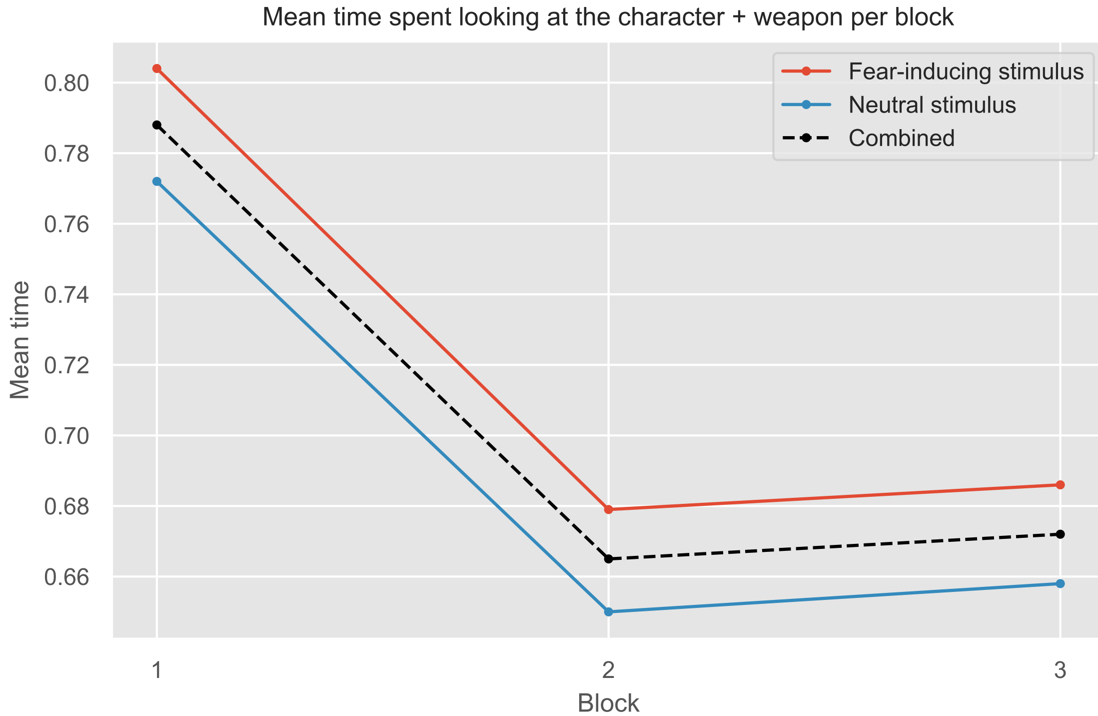


FIGURE 28: The average time spent looking at the character + weapon per stimulus type and block.

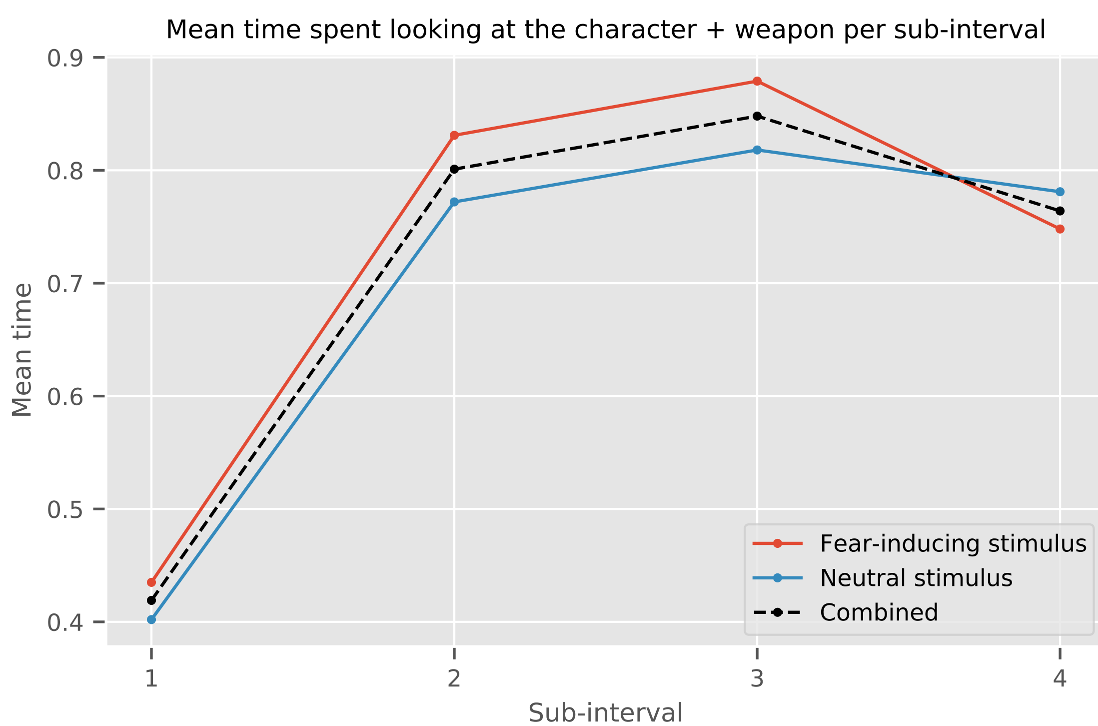


FIGURE 29: The average time spent looking at the character + weapon per stimulus type and sub-interval of blocks 1-3 combined.

effect for stimulus type, for the interaction block × sub-interval ($F(3.149, 44.085) = 1.225$, $p = .313$), for the interaction block × stimulus type ($F(2, 28) = .008$, $p = .992$), for the interaction sub-interval × stimulus type ($F(1.406, 19.686) = 1.005$, $p = .357$), and for the interaction block × sub-interval × stimulus type ($F(6, 84) = .738$, $p = .621$). Sphericity was not assumed for the effect block × sub-interval ($\chi^2(20) = 46.057, p = .001$) and for the effect sub-interval × stimulus type ($\chi^2(5) = 25.921, p < .001$), therefore the degrees of freedom were Greenhouse-Geisser corrected.

When looking at the average values in Figure 28, it appears that more time is spent looking at the stimuli during block one compared to blocks two and three, suggesting participants were looking slightly more at the environment during the following two blocks. This could be motivated by the participants needing more time to locate and perceive the stimuli, as they could enter the room while being outside of the participant's FOV. Blocks two and three appear to have minimal differences between each other's mean values, suggesting the uncertainty added by picking doors randomly instead of using fixed doors did not appear to influence overall gaze behavior majorly. Since the average difference between block one and blocks three and four is not very large, the decision was made not to analyze these further. The statistical difference for sub-interval could again be explained by the stimuli taking up more space within FOV as they approach the participant, so this significant effect does not appear to be very interesting either to analyze further.

4.5 Gaze Behavior Both Stimuli Simultaneously

block four contains both neutral and fear-inducing stimuli, which enter the room simultaneously. To determine which stimuli the gaze appeared to fixate on, the time spent looking at the individual characters and the weapon was analyzed. In Figure 40, the gaze targets per sub-interval and participant are visualized for block four. Here, the gaze targets are grouped per type of stimulus instead of categories for body and head separately. The weapon has its separate gaze target category, making it easier to determine if the gaze is directed towards the fear-inducing character or the weapon.

Character First, Statistical tests were done on the time spent looking at the individual characters, excluding the weapon. The average time spent looking at the characters can be seen in Figure 30.

An ANOVA on the mean time spent looking at the individual characters yielded a significant effect for interval ($F(1, 14) = 67.709$, $p < .001$). A nonsignificant effect was found for stimulus type ($F(1, 14) = .645$, $p = .435$), and the interaction interval × stimulus type ($F(1, 14) = .169$, $p = .687$). Interval appears to be significant for the same reason as before. When both characters stand in front of the participant, they take up a large portion of the FOV, compared to when they have yet to enter the room. From these result, it appears that none of the characters is more often looked at, which is also visible in Figure 40 by the almost even distribution of purple and red bars.

Character & Weapon The same statistical tests are done, including the weapon for the fear-inducing character, to determine if gaze again tends to centralize around the weapon. The average time spent looking at the characters and weapon combined can be seen in Figure 31.

An ANOVA on the mean time spent looking at the individual characters and weapon yielded a significant effect for interval ($F(1, 14) = 87.317$, $p < .001$). A significant effect is now found for stimulus type ($F(1, 14) = 11.362$, $p = .005$). An ANOVA on the interaction interval × stimulus type also appears to show

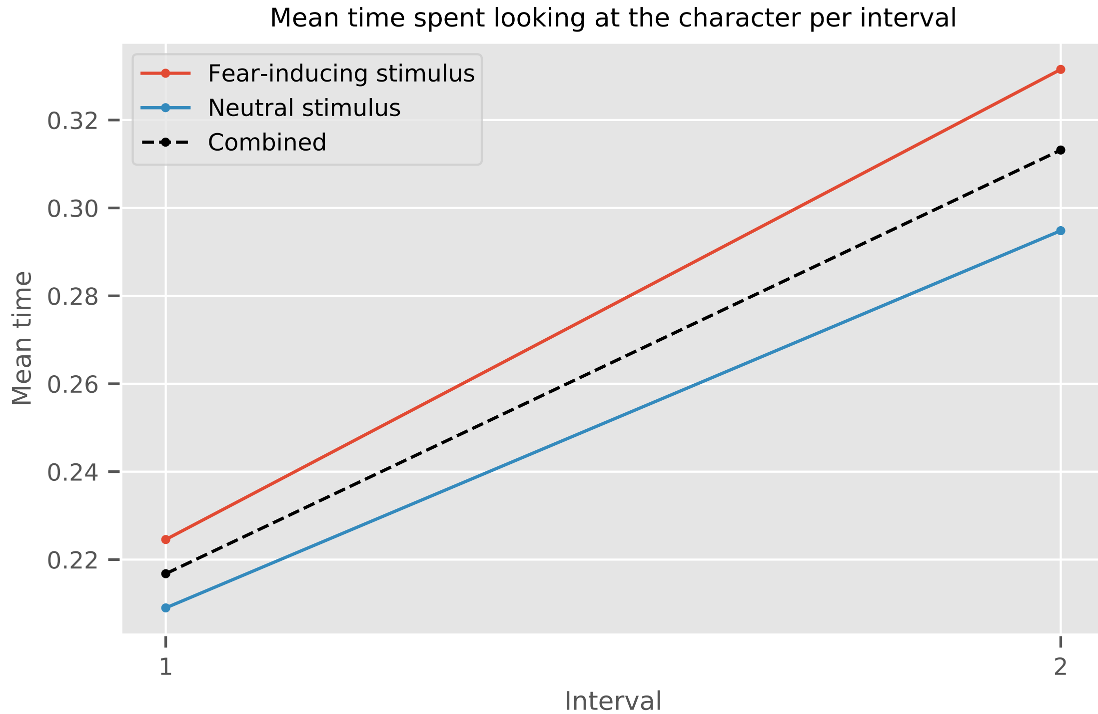


FIGURE 30: The average time spent looking at the character per stimulus type during block four.

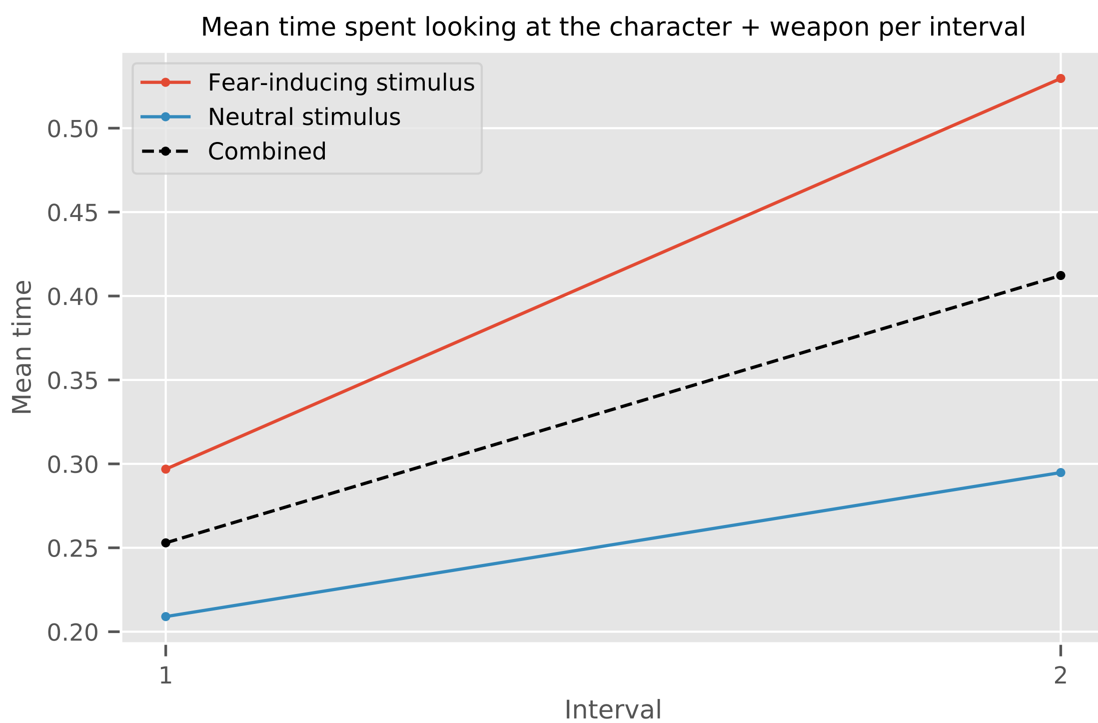


FIGURE 31: The average time spent looking at the character + weapon per stimulus type during block four.

a significant effect ($F(1, 14) = 5.958$, $p = .029$). Interval appears to be significant for the same reason as before.

When looking at Figure 31, it can be seen that the fear-inducing character and weapon are more looked at than the neutral character on average.

The interaction effect for interval × stimulus could be motivated by the average time spent looking at the fear-inducing stimulus and the weapon increasing stronger than the neutral stimulus during the second interval. This stronger increase could be motivated by the fear-inducing stimulus again taking up more space within view due to the weapon. Therefore the average time spent looking at this stimulus increases stronger during the second interval compared to the neutral stimulus.

From these results, it can be seen that individual characters are not looked at more often than one another. However, when the weapon is included, the difference is statistically significant, where the fear-inducing stimulus is more often looked at than the neutral stimulus. These two findings suggest that gaze tends to centralize towards the weapon when both stimuli were shown simultaneously.

Gaze Patterns over Time In Figure 40, the gaze targets per interval and participant are visualized for block four. Note that in this figure, the two intervals are again split up into four intervals.

It can be seen that again, in general, the time spent looking at the environment decreases over time. However, some participants stood out, who appeared to look relatively often at the environment throughout the entire duration of the trials, such as P06, P10, and P12. P12 even looks for a longer period at the environment than the stimuli during the last interval. When looking at their vocal fear-intensity ratings of this block, it appears that P10 did not appear to be afraid during all of the blocks (constantly gave zeroes), which might explain why this participant's gaze was less centralized towards the fear-inducing stimuli. P06 and P12 did appear to show larger intensities of fear during block four (scores within the range of 2-4), which might suggest this was evasive behavior.

P09 appeared to almost completely ignore the weapon. This participant gave a relatively high fear-intensity score of four to the fear-inducing stimulus and a zero to the neutral stimulus. This participant also appeared to look more often at the neutral stimulus, which might also suggest evasive behavior.

P13 also appeared to look more often at the neutral stimulus but gave both stimuli the same fear intensity score of three, which suggests this participant was in equal quantities afraid of both characters.

Overall, it appears that participants looked more often at the fear-inducing stimuli, even when not in fear (such as P08 and P14). It depends per participant if the weapon or the character is more looked at.

4.5.1 Discussion of Gaze Behavior

In Section 4.4.1, various analyses were performed on participants' gaze behavior. Results from Section 4.4.3 suggest that the average time spent looking at both types of stimuli was equal for blocks one, two, and three. However, the time spent looking at certain groups of gaze targets appeared to be different per stimulus type. Participants appeared to look more often at the neutral stimulus' face than the fear-inducing stimulus' face. Participants appeared to look more often at the weapon and face combined during the fear-inducing trials compared to the neutral trials, suggesting gaze centralized for a substantial portion towards the weapon.

Additionally, it appeared that the VE setting (e.g., one or six doors, fixed or randomly chosen doors) did not appear to influence the gaze behavior greatly, except slightly more time was required to perceive the stimuli during blocks two and three. When looking at the analyses of block four in Section 4.5, when both types of stimuli were presented, none of the characters was generally more often looked at when the weapon was excluded. When this was analyzed including the weapon in Section 4.5, again it appears that the gaze was mostly drawn towards the weapon. These results suggest that it did not matter whether one or two types of stimuli were presented. Gaze generally appeared to be drawn towards the fear-inducing stimuli. To the largest extent, this was the weapon, and to a smaller extent, this was the fear-inducing character's face.

In Section 4.4.1 and Section 4.5, gaze patterns were also analyzed per participant over time for blocks one and four, respectively, to determine whether more temporal gaze patterns would occur and whether differences occurred between participants. Results of block one showed that gaze behavior was more exploratory during the neutral trials, as more time was spent looking at the environment and character's body compared to the fear-inducing trials. During the fear-inducing trials, it appears that again, the attention is focused on the face and weapon, and to a much smaller extent to the body. Although participants generally followed these patterns, there were some differences between individuals. During the neutral trials, the average times spent fixating on the head, body, and environment appeared to be personal, as some participants completely fixated on the face and others spent more time fixating on the body. Some almost completely ignored the environment. For the fear-inducing trials, some participants looked more often at the weapon, and others looked more often at the face of the stimulus. A single participant appeared to look away from the weapon and face during the fear-inducing trials, which deviates from the general behavior patterns found, which suggests evasive behavior is occurring. Participants who claimed not to be afraid appeared to also follow the generally found gaze behavior during the fear-inducing trials, suggesting this behavior was also occurring subconsciously. Results of block four also found similar differences between participants in gaze behavior over time. Some participants ignored the characters and completely fixated on the weapon, while others appeared to ignore the weapon and focus on the characters instead. Some participants who claimed to feel high intensities of fear also appeared to look away from the fear-inducing stimuli, which again suggested evasive behavior. These findings were in line with related literature from Section 2.2.3, which discussed that people allocate more time to arousing stimuli or threatening cues and centralize their gaze more towards the fear-inducing stimuli. Additionally, in this section it was discussed that the gaze centralizing around fear-inducing stimuli are related to freezing-like patterns.

When the results of block four are compared to results of blocks one, two, and three, there generally appear to be no major differences in behavior if one type of stimuli is presented versus both types simultaneously. However, vocal feedback suggested that some participants appeared to feel higher intensities of fear when approached by two characters during block four, as discussed in Section 4.1.5. Participants felt intimidated by the characters, so some participants appeared to try to evade them, especially when both characters got relatively high or equal fear-intensity scores. This evasive behavior occasionally found in participants appears to be in line with the characteristics of fear behavior, as fear is generally described as threat-avoiding behavior, as discussed by related works in Section 2.3.2.

Section 4.4.2 analyzed pupil dilation behavior. However, this behavior did not appear to change significantly based on stimulus type, which is not in line with the relations found between arousing stimuli and pupil dilation in related works, as discussed in Section 2.2.3. This might be motivated by the usage of a VR device. The participant has displays right in front of the eyes, which might show adjusted pupil dilation behavior, as the lighting conditions are different compared to real-world lighting conditions.

4.6 Delay of Adjusted Behavior

In the previous sections it was found that generally behavior appears to deviate during the fear-inducing trials based on some of the measurements, which suggests fear behavior was induced in the participants. To determine whether this behavior would manifest with a delay, further analyses were done on the fear-inducing trials. For these trials, the body sway and balance data were again split up into four sub-intervals, as done in earlier analyses. Each main interval was statistically compared to its second sub-interval. For instance, the behavior during the first interval (in which the stimulus enters the room and approaches the participant) is compared to the last fraction of this interval (where the stimulus has already entered through the door and is nearing the participant). This is also done for the second interval (where the stimulus stands still and fires the rifle), where it is compared to last fraction (when the gunshot had taken place). These comparisons were done to determine whether the last sub-intervals appear to show significantly deviating behavior compared to the general behavior of the main interval, meaning the moments in time could be found where fear-behavior would be occurring.

4.6.1 Delay in Body Sway Adjustments

The following analyses of the delay in body sway were done on the fear-inducing trials of block one. During these analyses, the first interval will be referred to as sub-intervals 1+2, and the second interval will be referred to as sub-intervals 3+4.

An ANOVA on the mean Euclidean distance yielded a significant difference between sub-intervals 1+2 and sub-interval 2 ($F(1, 15) = 4.567$, $p = .049$). An ANOVA on the mean absolute distance in ML-direction yielded a marginal difference between interval 1+2 and interval 2 ($F(1, 15) = 4.108$, $p = .061$). An ANOVA on the mean absolute distance in AP-direction yielded a significant difference between interval 1+2 and interval 2 ($F(1, 15) = 4.790$, $p = .045$). The average body sway measurement values of these sub-intervals are visualized in Figure 32.

An ANOVA on the mean Euclidean distance did not yield a significant difference between sub-intervals 3+4 and sub-interval 4 ($F(1, 15) = 2.158$, $p = .163$). An ANOVA on the mean absolute distance in ML-direction did not yield a significant difference between sub-intervals 3+4 and sub-interval 4 ($F(1, 15) = .033$, $p = .858$). An ANOVA on the mean absolute distance in AP-direction yielded a significant difference between sub-intervals 3+4 and sub-interval 4 ($F(1, 15) = 4.915$, $p = .042$). The average body sway measurement values of these sub-intervals are visualized in Figure 33.

These results overall suggest that body sway behavior appeared to deviate in sub-interval 2, as the body sway measurements were all higher on average during this sub-interval than during the sub-intervals 1+2 combined, as can be seen in Figure 32. This might suggest that fear behavior started to show up as the character got closer to the participant. The results of sub-intervals 3+4 only show significant differences in the AP-direction, which might be explained by the participants startling due to the rifle shot. However, when looking at Figure 33, this difference appears to be relatively small. As the other two measurements appear not to show significant differences, it appears that behavior had already been adjusted before these sub-intervals. From the balance data, it appears that this adjusted behavior manifested with a delay. This delay appeared to last until the stimulus was approaching the participant. When looking at Figure 22 from earlier analyses, it appeared that the body sway measurements also appeared to be almost equal for both types of stimuli during

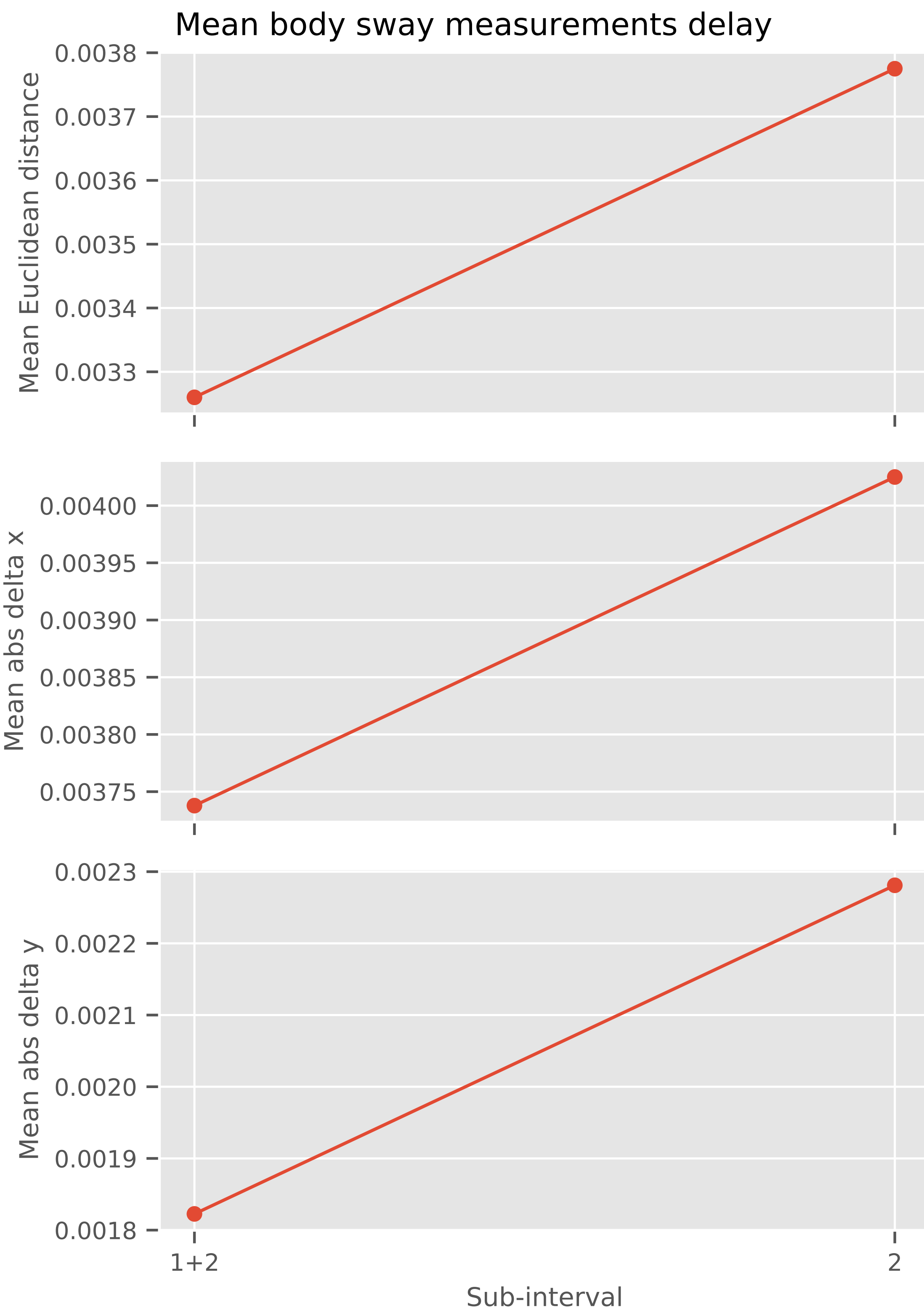


FIGURE 32: Mean body sway measurement values of interval 1 (sub-intervals 1+2) compared to its last sub-interval 2.

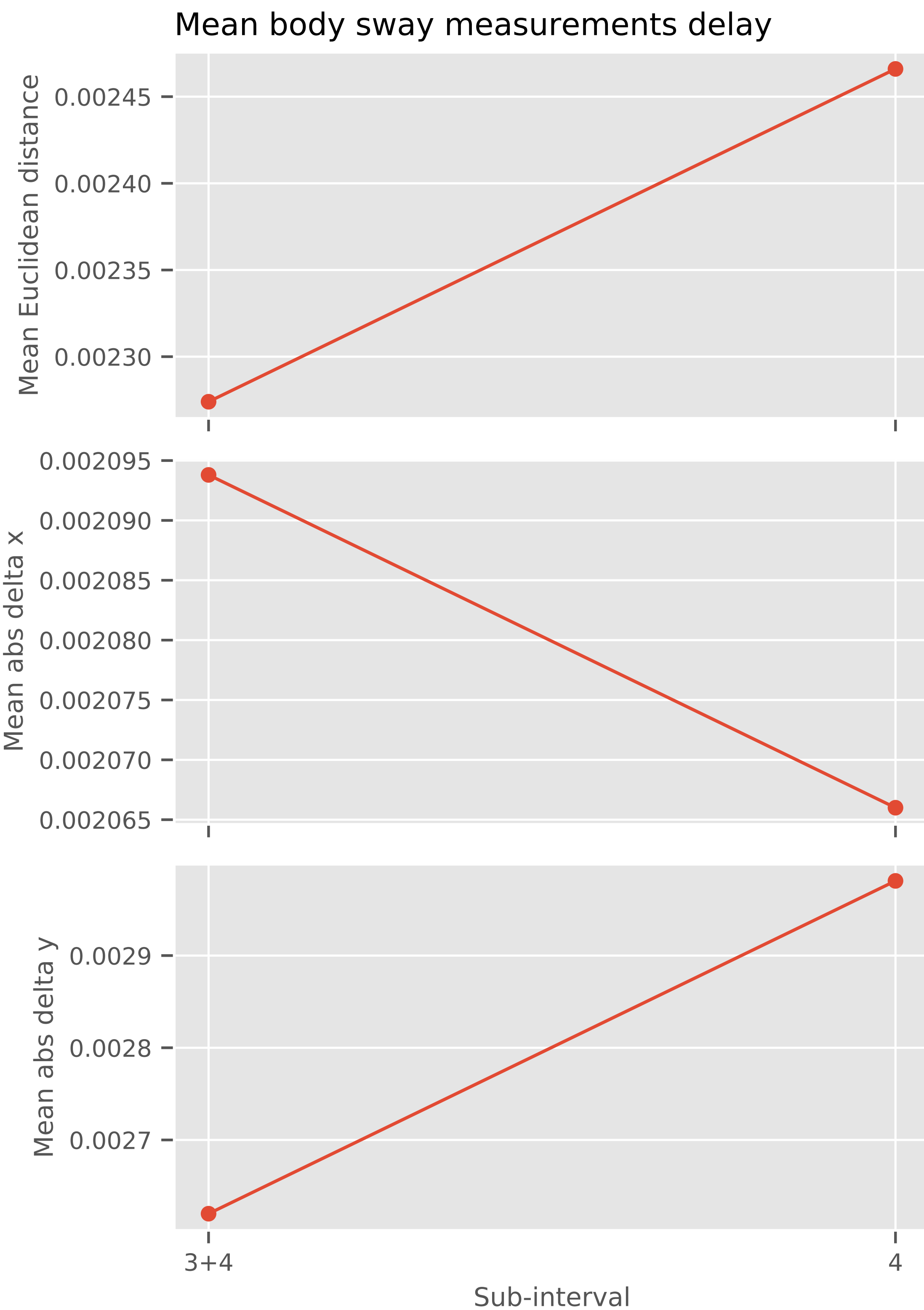


FIGURE 33: Mean body sway measurement values of interval 2 (sub-intervals 3+4) compared to its last sub-interval 4.

sub-interval one, while these values appeared to increase starting from sub-interval two, again suggesting the delay of behavior ends somewhere between these two sub-intervals in time.

4.6.2 Delay in Gaze Behavior Adjustments

The following analyses of the delay in gaze behavior were also done on the fear-inducing trials of block one. Here, the time spent looking at the important stimuli elements, the head, and weapon, will be analyzed. The average times spent looking at the head and weapon combined are visualized in Figure 34.

An ANOVA on the mean time spent looking at the head and weapon combined yielded a significant difference between sub-intervals 1+2 and sub-interval 2 ($F(1, 14) = 72.074$, $p < .001$). An ANOVA on the mean time spent looking at the head and weapon combined yielded a marginal difference between sub-intervals 3+4 and sub-interval 4 ($F(1, 14) = 3.540$, $p = .081$).

The marginal difference between sub-intervals 3+4 and sub-interval 4 and the average values of Figure 34 suggest that the participants look less often at the stimulus during sub-interval 4, which does not suggest fear behavior started occurring from this point in time. Additionally, the difference between the averages is relatively small. Therefore, fear could possibly already be induced before sub-intervals 3 and 4, which again points in the direction of sub-intervals 1 and 2. The significant difference between sub-intervals 1+2 and sub-interval 2 could again be explained by the larger space the character takes up in the FOV as it nears the participants. However, it could also partly be explained by adjusted behavior occurring at this point in time. Also looking at the differences between sub-interval 2 and sub-intervals 3+4, there appears to be a relatively large difference, which might suggest that participants started looking more often at the stimuli when it had approached them and stood still in front of the participants. When looking at Figure 29, it also appears that for blocks one, two, and three combined the differences between the time spent looking at the fear-inducing stimuli and neutral stimuli are much smaller during sub-interval one compared to the other sub-intervals, which give the same suggestion as the other findings that a delay occurred which ended somewhere between these sub-intervals.

4.6.3 Discussion of Delay

The results of gaze and body sway combined suggest there appears to be a delay when adjusted behavior showed up during the fear-inducing trials. Results point at sub-intervals 1 and 2. However, it is unclear when exactly fear behavior started occurring in the gaze behavior as the differences in behavior could also be caused by the space the stimuli took up in the FOV. Including the body sway measurements, fear behavior might be seen as the character starts nearing the participant or stands in front of the participant.

FIGURE 34: Mean body sway measurements per interval compared to its last sub-interval. On the left, interval 1 (sub-intervals 1+2) is compared to sub-interval 2. On the right, interval 2 (sub-intervals 3+4) is compared to sub-interval 4.

5 Conclusion

In this thesis, various analyses were performed on the body sway and gaze behavior data of participants, who were approached by fear-inducing and neutral stimuli in VR. With the results obtained in Chapter 4, the research questions of Section 1.6 will be answered. Sections 5.1, 5.2, and 5.3 will answer the three research questions, respectively. Section 5.4 will discuss limitations of the proposed method and will lay the foundation for future research. Finally, Section 5.5 will conclude the thesis by assessing the proposed method's ability to analyze fear and freezing behavior based on the answers to the research questions.

5.1 Measured Effect in Body Motion and Gaze Behavior

The first research question was posed to analyze the differences in measured effect in body motion and gaze behavior between fear-inducing and neutral stimuli. Body sway results suggested that there is little evidence for a generally shared difference in body sway between the types of stimuli presented. A difference is found for the ML-direction, however evidence of this is not very strong. When looking at an individual level, the results were divided into two types of behavior. Some participants appeared to show increased body sway and other participants appeared to show reduced body sway when the fear-inducing stimuli were presented compared to when neutral stimuli were presented. This latter group of participants showed the strongest evidence that freezing behavior occurred when looking at their body sway. It is unclear why the participants show opposing body sway behavior, so this should be further investigated in future research. Therefore, based on these results it cannot be said with confidence that body sway was a good measure for detecting fear behavior.

Pupil size did not appear to change based on the stimulus type presented. Therefore, pupil size did not appear to be a good measure for detecting fear behavior.

Gaze fixation behavior results suggested that gaze appeared to change based on the type of stimuli presented. During the fear-inducing trials, gaze appeared to be drawn more towards the fear-inducing stimuli, which was to the largest extent the weapon, while gaze appeared to be more exploratory during the neutral trials. Additionally, it was found that this adjusted behavior appeared to be subconscious. These results were in line with related literature about fear behavior, so gaze fixation behavior measurements showed to be of great use in the detection of fear behavior.

When combining these results, it appeared that there were indeed differences in measured effect between when fear-inducing stimuli were presented compared to neutral stimuli for the measurements of body sway and gaze fixations. It appeared that overall the participants showed correlates of fear in their body motion and gaze behavior when confronted by the fear-inducing stimuli. Occasionally, correlates of freezing behavior were shown in the body sway behavior of participants and the threat-directed gaze behavior also appears to show freezing-like behavior.

5.2 Gaze Centralization

The second research question was posed to analyze how gaze behavior would be adjusted when both types of stimuli were presented simultaneously, and if any of the stimuli would draw more attention. It was generally found that gaze appeared to centralize more towards the fear-inducing stimuli, which was in line with related literature about fear and freezing behavior. When looking at an individual level, there appear to be some subtle differences in the distribution of time spent looking at the gaze targets, however, most participants followed the general pattern found, which was again that a large portion of attention was spent fixating on the weapon. Additionally, it was found that some participants who experienced high intensities of fear showed evasive behavior. Here, it was also found that the gaze behavior appeared to be subconscious, based on subjective fear-intensity scores. Therefore, it can also be concluded that fear and freezing behavior can be found when fear-inducing and neutral stimuli are presented simultaneously when gaze fixation behavior is measured.

5.3 Delay in Fear Behavior

The final research question was posed to analyze whether the found effects in body motion and gaze behavior would manifest with a delay. Both body sway and gaze behavior appeared to be adjusted at the moments the fear-inducing characters were nearing or standing still in front of the participants. Adjusted behavior was not found yet when the door would open and the character would still walk relatively far away from the participant, suggesting the moment when fear-intensity increases could be related to the distance the participant stood from the stimuli. No exact moment in time could be determined at which the delay had ended. Due to other factors, such as the size of the stimuli in the FOV and the conflicting results of fear behavior in body sway, it could not be determined with confidence from the results that a delay occurred before fear behavior appeared to manifest. Therefore it would provide useful to analyze this effect delay in future research.

5.4 Limitations & Future Work

When this thesis was carried out, some limitations and inspirations for future work came to light, which will be discussed in more detail in this section.

5.4.1 Participant Recruitment

Ideally, the number of participants should have been higher, as 19 was slightly on the lower side, which left little room for errors or invalid sessions. Unfortunately, three sessions' sensor data could not be used, and one session's gaze data could not be used, resulting in only 15 successfully captured sessions. Therefore, interesting or deviating patterns in the sensor data were mostly found in only one or two participants, so these could not be further inspected. Additionally, a stricter recruitment process should have been used, as participants who wore lenses were able to participate in the experiment. It is unclear to what extent lenses could have negatively influenced gaze tracking performance due to the refraction of light. However, an even smaller number of participants would have been recruited if this restriction would have also been in place.

5.4.2 Subjective Measurements

There were some inconsistencies in the scoring of subjective measurements throughout the experiment. Most questionnaire items about the stimuli used a 1-5 Likert scale, while the vocal answers in-between blocks were in the range of 0-5. Additionally, some participants unexpectedly gave decimal answers in-between the blocks, which were supposed to be integer values. These made it challenging to compare scores assigned to the stimuli in-between blocks and their final verdict after the session. Therefore, in future works, the scoring of subjective measurements should be done more consistently and follow the same rules. Additionally, it could provide useful to include the STAI and TIS scores directly in the statistical analyses in future works, which was not done in this thesis due to time restrictions. By doing this, behavior could be better linked to personal fear intensities.

5.4.3 Gaze Tracking Solution

The built-in Tobii eye tracker gave limited insights into gaze calibration quality, which made it challenging to assess the quality of the gaze data. Secondly, this gaze tracker did not clearly label whether a loss of signal was caused by a blink or by the tracker malfunctioning. These made it difficult to define other interesting gaze behavior measurements, such as gaze fixations and blinking rate. Therefore, it might provide useful to further analyze the quality of this gaze tracker or to use a different gaze tracking solution in future works. Additionally, it could provide useful to use the blinking rate in future works, as this could give more insight into fear and freezing behavior.

5.4.4 Virtual Environment & Characters

The used stimuli appeared to be of reasonable quality for the experiments, as they were able to induce more fear as their fear-inducing variants compared to their neutral variants. However, it appeared that character model Josh appeared to still induce some subtle intensities of fear as a neutral stimulus. Additionally, participant ratings of the realism of the characters and VE were not very high, suggesting participants were not impressed with the way the elements of the VE looked, which might have negatively influenced immersion. Therefore, in future works, more time should be dedicated to selecting high-quality stimuli and to designing a closer to realism VE, which might benefit immersion and could induce more natural behavior.

5.4.5 Data from Other Modalities

Further research should also include motion capture, which unfortunately had to be left out during this research due to hardware-related issues. Related works suggest posture and joint kinematics also play a role in fear behavior, so including joint position and orientation data could be beneficial for the recognition of fear behavior. Additionally, it could provide useful to include vocal behavior, as a relation between vocal pitch and fear was found in related literature. Body temperature and heart rate are other options for measurements, though these are more challenging to include unobtrusively.

5.4.6 Body Sway Behavior

Further research should be dedicated to why the induced fear behavior appears to manifest in both more and less body movement. It appeared that the participants could be divided into two groups with opposing patterns of behavior. No clear cause could be found why this was the case, so therefore it might be interesting to analyze this further. Additionally, in general, the body sway in ML-direction appeared to show an increase in movement during the fear-inducing trials. This was unexpected, as the correlates of fear suggest that body sway should be reduced instead. This increase in body sway could not be explained, so it would be interesting to further analyze its cause.

5.4.7 Virtual Reality Behavioral Studies

The proposed method showed evidence that natural behavior can be induced in VR, and its use could therefore be beneficial for other behavioral studies. For instance, further studies could be performed analyzing other types of fear stimuli. Studies should also be conducted analyzing other types of emotional behavior besides fear.

5.4.8 Machine Learning Applications

With the aid of machine learning techniques, future research could aid in the automatic recognition of freezing behavior. Results have shown that some temporal or subtle behavior patterns occur over time, for instance in the body sway measurements. Although most of the peaks in the signals could be explained by events occurring during the trials or by coincidental movements, perhaps more subtle or hidden patterns could be found or classified over time by using deep learning techniques, such as LSTMs.

5.5 Conclusion Proposed Method

This thesis project has introduced a novel multimodal approach to recognizing freezing behavior using VR. As evidence of fear behavior was found in the results of both the sensor data and the questionnaire answers, it appears that the usage of VR was justified and may have contributed to showing natural behavior in participants while not being as intrusive as other sensor- or marker-based approaches. Additionally, the proposed method appeared to be beneficial in the induction and recognition of fear behavior despite its relatively low costs. Furthermore, it even provided results that showed freezing behavior on some occasions. Therefore, it can be concluded that the main goal of improving the present ability to recognize freezing behavior in humans unobtrusively using multiple modalities has been accomplished.

References

[1] Robert J Blanchard, Kevin J Flannelly, and D Caroline Blanchard. Defensive behaviors of laboratory and wild rattus norvegicus.. *Journal of Comparative Psychology*, 100(2):101, 1986.

[2] Karin Roelofs, Muriel A Hagenaars, and John Stins. Facing freeze: social threat induces bodily freeze in humans. *Psychological science*, 21(11):1575–1581, 2010.

[3] Muriel A Hagenaars, John F Stins, and Karin Roelofs. Aversive life events enhance human freezing responses. *Journal of Experimental Psychology: General*, 141(1):98, 2012.

[4] Muriel A Hagenaars, Melly Oitzl, and Karin Roelofs. Updating freeze: aligning animal and human research. *Neuroscience & Biobehavioral Reviews*, 47:165–176, 2014.

[5] Byron A Campbell, Gwendolyn Wood, and Thomas McBride. Origins of orienting and defensive responses: An evolutionary perspective. *Attention and orienting: Sensory and motivational processes*, pages 41–67, 1997.

[6] Marie-Claire Chelini, Rodrigo H Willemart, and Eileen A Hebets. Costs and benefits of freezing behaviour in the harvestman eumesosoma roeweri (arachnida, opiliones). *Behavioural Processes*, 82(2):153–159, 2009.

[7] Ronald Poppe, Sophie Van Der Zee, Dirk KJ Heylen, and Paul J Taylor. Amab: Automated measurement and analysis of body motion. *Behavior research methods*, 46(3):625–633, 2014.

[8] Alessandro Vinciarelli, Maja Pantic, and Hervé Bourlard. Social signal processing: Survey of an emerging domain. *Image and vision computing*, 27(12):1743–1759, 2009.

[9] Jong-Suk Choi, Jae Won Bang, Hwan Heo, and Kang Ryoung Park. Evaluation of fear using nonintrusive measurement of multimodal sensors. *Sensors*, 15(7):17507–17533, 2015.

[10] Iain R Murray and John L Arnott. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America*, 93(2):1097–1108, 1993.

[11] Paul Ekman. Emotions revealed. *Bmj*, 328(Suppl S5), 2004.

[12] Viviane Clay, Peter König, and Sabine U König. Eye tracking in virtual reality. *Journal of Eye Movement Research*, 12(1), 2019.

[13] Rosenberg Ekman. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA, 1997.

[14] James A Russell. Is there universal recognition of emotion from facial expression? a review of the cross-cultural studies. *Psychological bulletin*, 115(1):102, 1994.

[15] MW Schurgin, J Nelson, S Iida, H Ohira, JY Chiao, and SL Franconeri. Eye movements during emotion recognition in faces. *Journal of vision*, 14(13):14–14, 2014.

[16] Paul Ekman. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200, 1992.

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

[18] James A Russell and Merry Bullock. Fuzzy concepts and the perception of emotion in facial expressions. *Social Cognition*, 4(3):309–341, 1986.

[19] Robert Plutchik. *The emotions*. University Press of America, 1991.

[20] Robert Plutchik. A general psychoevolutionary theory of emotion. In *Theories of emotion*, pages 3–33. Elsevier, 1980.

[21] Machine Elf 1735. Robert plutchik's wheel of emotions. https://en.wikipedia.org/wiki/File:Plutchik-wheel.svg, 2011. Accessed: February 3, 2021.

[22] Friesen W Ekman P. Facial action coding system: A technique for the measurement of facial movement. *Palo Alto: Consulting Psychologists Press*, 1978. Manual cannot be accessed anymore and is now being continued as an automated system, which can be accessed via the Paul Ekman Group website. Accessed: January 13, 2021.

[23] Lawrence Ian Reed, Michael A Sayette, and Jeffrey F Cohn. Impact of depression on response to comedy: A dynamic facial coding analysis. *Journal of abnormal psychology*, 116(4):804, 2007.

[24] Ruicong Zhi, Mengyi Liu, and Dezheng Zhang. A comprehensive survey on automatic facial action unit analysis. *The Visual Computer*, 36(5):1067–1093, 2020.

[25] Peter J Lang, Margaret M Bradley, Bruce N Cuthbert, et al. Motivated attention: Affect, activation, and action. *Attention and orienting: Sensory and motivational processes*, 97:135, 1997.

[26] Tal Shafir, Rachelle P Tsachor, and Kathleen B Welch. Emotion regulation through movement: unique sets of movement characteristics are associated with and enhance basic emotions. *Frontiers in psychology*, 6:2030, 2016.

[27] Mariska Esther Kret, Karin Roelofs, Jeroen Stekelenburg, and Beatrice de Gelder. Emotional signals from faces, bodies and scenes influence observers' face expressions, fixations and pupil-size. *Frontiers in human neuroscience*, 7:810, 2013.

[28] Lisette J Schmidt, Artem V Belopolsky, and Jan Theeuwes. Potential threat attracts attention and interferes with voluntary saccades. *Emotion*, 15(3):329, 2015.

[29] Alma-Sophia Merscher and Matthias Gamer. Freezing of gaze in anticipation of avoidable threat: A threat-specific proxy for freezing-like behavior in humans. In *ACM Symposium on Eye Tracking Research and Applications*, pages 1–3, 2020.

[30] Arne Öhman and Susan Mineka. Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. *Psychological review*, 108(3):483, 2001.

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

[32] Timo Partala and Veikko Surakka. Pupil size variation as an indication of affective processing. *International journal of human-computer studies*, 59(1-2):185–198, 2003.

[33] Henk Van Steenbergen, Guido PH Band, and Bernhard Hommel. Threat but not arousal narrows attention: evidence from pupil dilation and saccade control. *Frontiers in psychology*, 2:281, 2011.

[34] Erno J Hermans, Marloes JAG Henckens, Karin Roelofs, and Guillén Fernández. Fear bradycardia and activation of the human periaqueductal grey. *Neuroimage*, 66:278–287, 2013.

[35] Wei-Long Zheng, Bo-Nan Dong, and Bao-Liang Lu. Multimodal emotion recognition using eeg and eye tracking data. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5040–5043. IEEE, 2014.

[36] Claudio Aracena, Sebastián Basterrech, Václav Snáel, and Juan Velásquez. Neural networks for emotion recognition based on eye tracking data. In *2015 IEEE International Conference on Systems, Man, and Cybernetics*, pages 2632–2637. IEEE, 2015.

[37] Christian Grillon, Rezvan Ameli, Scott W Woods, Kathleen Merikangas, and Michael Davis. Fear-potentiated startle in humans: Effects of anticipatory anxiety on the acoustic blink reflex. *Psychophysiology*, 28(5):588–595, 1991.

[38] Stephen Maren. Neurobiology of pavlovian fear conditioning. *Annual review of neuroscience*, 24(1):897–931, 2001.

[39] Lara Rösler and Matthias Gamer. Freezing as flight preparation: Evidence from electrodermal, cardiovascular and gaze recordings. 2019.

[40] Surya Cheemalapati, Michael Gubanov, Michael Del Vale, and Anna Pyayt. A real-time classification algorithm for emotion detection using portable eeg. In *2013 IEEE 14th International Conference on Information Reuse & Integration (IRI)*, pages 720–723. IEEE, 2013.

[41] Alex Pentland. *Honest signals: how they shape our world*. MIT press, 2010.

[42] Haris Zacharatos, Christos Gatzoulis, and Yiorgos L Chrysanthou. Automatic emotion recognition based on body movement analysis: a survey. *IEEE computer graphics and applications*, 34(6):35–45, 2014.

[43] Charles Darwin. *The Expression of the Emotions in Man and Animals*. University of Chicago Press, Chicago, 1872.

[44] Daniel Bernhardt. *Emotion inference from human body motion*. PhD thesis, Citeseer, 2010.

[45] Frank E Pollick, Helena M Paterson, Armin Bruderlin, and Anthony J Sanford. Perceiving affect from arm movement. *Cognition*, 82(2):B51–B61, 2001.

[46] Roddy Cowie, Ellen Douglas-Cowie, Nicolas Tsapatsoulis, George Votsis, Stefanos Kollias, Winfried Fellenz, and John G Taylor. Emotion recognition in human-computer interaction. *IEEE Signal processing magazine*, 18(1):32–80, 2001.

[47] Suk Kyu Lee, Mungyu Bae, Woonghee Lee, and Hwangnam Kim. Cepp: Perceiving the emotional state of the user based on body posture. *Applied Sciences*, 7(10):978, 2017.

[48] Tomasz Sapiński, Dorota Kamińska, Adam Pelikant, and Gholamreza Anbarjafari. Emotion recognition from skeletal movements. *Entropy*, 21(7):646, 2019.

[49] Kyriaki Kaza, Athanasios Psaltis, Kiriakos Stefanidis, Konstantinos C Apostolakis, Spyridon Thermos, Kosmas Dimitropoulos, and Petros Daras. Body motion analysis for emotion recognition in serious games. In *International Conference on Universal Access in Human-Computer Interaction*, pages 33–42. Springer, 2016.

[50] Ross A Clark, Adam L Bryant, Yonghao Pua, Paul McCrory, Kim Bennell, and Michael Hunt. Validity and reliability of the nintendo wii balance board for assessment of standing balance. *Gait & posture*, 31(3):307–310, 2010.

[51] Gerwin de Haan, Eric J Griffith, and Frits H Post. Using the wii balance board™ as a low-cost vr interaction device. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, pages 289–290, 2008.

[52] Robert E Schapire. Explaining adaboost. In *Empirical inference*, pages 37–52. Springer, 2013.

[53] Kenji Mase. Recognition of facial expression from optical flow. *IEICE TRANSACTIONS on Information and Systems*, 74(10):3474–3483, 1991.

[54] Yaser Yacoob and Larry S Davis. Recognizing human facial expressions from long image sequences using optical flow. *IEEE Transactions on pattern analysis and machine intelligence*, 18(6):636–642, 1996.

[55] Jeffrey F Cohn, Adena J Zlochower, James J Lien, and Takeo Kanade. Feature-point tracking by optical flow discriminates subtle differences in facial expression. In *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, pages 396–401. IEEE, 1998.

[56] Dave M Stampe. Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. *Behavior Research Methods, Instruments, & Computers*, 25(2):137–142, 1993.

[57] Sarah Uzzaman and Steve Joordens. The eyes know what you are thinking: eye movements as an objective measure of mind wandering. *Consciousness and cognition*, 20(4):1882–1886, 2011.

[58] Ramtin Zargari Marandi, Pascal Madeleine, Øyvind Omland, Nicolas Vuillerme, and Afshin Samani. Eye movement characteristics reflected fatigue development in both young and elderly individuals. *Scientific reports*, 8(1):1–10, 2018.

[59] Alexandra Sipatchin, Siegfried Wahl, and Katharina Rifai. Eye-tracking for low vision with virtual reality (vr): testing status quo usability of the htc vive pro eye. *bioRxiv*, 2020.

[60] Alexandra Sipatchin, Siegfried Wahl, and Katharina Rifai. Accuracy and precision of the htc vive pro eye tracking in head-restrained and head-free conditions. *Investigative Ophthalmology & Visual Science*, 61(7):5071–5071, 2020.

[61] Jae Won Bang, Hwan Heo, Jong-Suk Choi, and Kang Ryoung Park. Assessment of eye fatigue caused by 3d displays based on multimodal measurements. *Sensors*, 14(9):16467–16485, 2014.

[62] Kwang Yong Shin, Yeong Gon Kim, and Kang Ryoung Park. Enhanced iris recognition method based on multi-unit iris images. *Optical Engineering*, 52(4):047201, 2013.

[63] Jakob Lemos. System and method for determining human emotion by analyzing eye properties, March 22 2007. US Patent App. 11/522,476.

[64] HTC. Htc vive. https://www.vive.com/us/, 2021. Accessed: February 10, 2021.

[65] D Caroline Blanchard and Robert J Blanchard. Ethoexperimental approaches to the biology of emotion. *Annual review of psychology*, 39(1):43–68, 1988.

[66] Neil McNaughton and Philip J Corr. A two-dimensional neuropsychology of defense: fear/anxiety and defensive distance. *Neuroscience & Biobehavioral Reviews*, 28(3):285–305, 2004.

[67] Patrick Sylvers, Scott O Lilienfeld, and Jamie L LaPrairie. Differences between trait fear and trait anxiety: Implications for psychopathology. *Clinical psychology review*, 31(1):122–137, 2011.

[68] Phil Barker. *Psychiatric and mental health nursing: The craft of caring*. CRC Press, 2017.

[69] Grigore C Burdea and Philippe Coiffet. *Virtual reality technology*. John Wiley & Sons, 2003.

[70] JM Zheng, KW Chan, and Ian Gibson. Virtual reality. *Ieee Potentials*, 17(2):20–23, 1998.

[71] Diederick C Niehorster, Li Li, and Markus Lappe. The accuracy and precision of position and orientation tracking in the htc vive virtual reality system for scientific research. *i-Perception*, 8(3):2041669517708205, 2017.

[72] Hugh Langley. Inside-out v outside-in: How vr tracking works, and how it's going to change. *URL: https://www. wareable. com/vr/inside-out-vs-outside-in-vr-tracking-343*, pages 1–7, 2017.

[73] Daniel Pustka, Jan-Patrick Hülß, Jochen Willneff, Frieder Pankratz, Manuel Huber, and Gudrun Klinker. Optical outside-in tracking using unmodified mobile phones. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 81–89. IEEE, 2012.

[74] Sony. Playstation vr. https://www.playstation.com/en-us/ps-vr/, 2021. Accessed: February 10, 2021.

[75] Google. Google cardboard. https://vr.google.com/cardboard/, 2021. Accessed: February 10, 2021.

[76] Frank Nielsen. Surround video: a multihead camera approach. *The visual computer*, 21(1):92–103, 2005.

[77] Valve Corporation. Steamvr. https://www.steamvr.com, 2021. Accessed: February 10, 2021.

[78] Unity Technologies. Unity. https://unity.com/, 2021. Accessed: January 5, 2021.

[79] Epic Games. Unreal engine. https://www.unrealengine.com. Accessed: January 5, 2021.

[80] M Pavlou, RG Kanegaonkar, David Swapp, DE Bamiou, Mel Slater, and LM Luxon. The effect of virtual reality on visual vertigo symptoms in patients with peripheral vestibular dysfunction: a pilot study. *Journal of vestibular research*, 22(5, 6):273–281, 2012.

[81] Cristina Botella, Helena Villa, S Quero, RM Baños, M Alcaniz, et al. The use of vr in the treatment of panic disorders and agoraphobia. *Studies in health technology and informatics*, 99:73–90, 2004.

[82] Stéphane Bouchard, Sophie Côté, Julie St-Jacques, Geneviève Robillard, and Patrice Renaud. Effectiveness of virtual reality exposure in the treatment of arachnophobia using 3d games. *Technology and health care*, 14(1):19–27, 2006.

[83] Devon Allcoat, W James Greville, Philip M Newton, and Simon Dymond. Frozen with fear: conditioned suppression in a virtual reality model of human anxiety. *Behavioural processes*, 118:98–101, 2015.

[84] DE Blackman. Conditioned anxiety and operant behavior. *Schedule effects: Drugs, drinking, and aggression*, pages 26–49, 1972.

[85] Derek E Blackman. *Operant conditioning: an experimental analysis of behaviour*. Routledge, 2017.

[86] Jack Brookes, Matthew Warburton, Mshari Alghadier, Mark Mon-Williams, and Faisal Mushtaq. Studying human behavior with virtual reality: The unity experiment framework. *Behavior research methods*, pages 1–9, 2019.

[87] Bob G Witmer and Michael J Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3):225–240, 1998.

[88] Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993.

[89] Eugenia M Kolasinski. *Simulator sickness in virtual environments*, volume 1027. US Army Research Institute for the Behavioral and Social Sciences, 1995.

[90] Hyun K Kim, Jaehyun Park, Yeongcheol Choi, and Mungyeong Choe. Virtual reality sickness questionnaire (vrsq): Motion sickness measurement index in a virtual reality environment. *Applied ergonomics*, 69:66–73, 2018.

[91] Hironori Akiduki, Suetaka Nishiike, Hiroshi Watanabe, Katsunori Matsuoka, Takeshi Kubo, and Noriaki Takeda. Visual-vestibular conflict induced by virtual reality in humans. *Neuroscience letters*, 340(3):197–200, 2003.

[92] Hironori Akizuki, Atsuhiko Uno, Kouichi Arai, Soukichi Morioka, Seizo Ohyama, Suetaka Nishiike, Koichi Tamura, and Noriaki Takeda. Effects of immersion in virtual reality on postural control. *Neuroscience letters*, 379(1):23–26, 2005.

[93] Bruno Patrão, Samuel Pedro, and Paulo Menezes. How to deal with motion sickness in virtual reality. 2020.

[94] Justin Munafo, Meg Diedrick, and Thomas A Stoffregen. The virtual reality head-mounted display oculus rift induces motion sickness and is sexist in its effects. *Experimental brain research*, 235(3):889–901, 2017.

[95] A Lawther and MJ Griffin. A survey of the occurrence of motion sickness amongst passengers at sea. *Aviation, space, and environmental medicine*, 59(5):399–406, 1988.

[96] Doug A Bowman, David Koller, and Larry F Hodges. Travel in immersive virtual environments: An evaluation of viewpoint motion control techniques. In *Proceedings of IEEE 1997 Annual International Symposium on Virtual Reality*, pages 45–52. IEEE, 1997.

[97] Jiwan Bhandari, Paul R MacNeilage, and Eelke Folmer. Teleportation without spatial disorientation using optical flow cues. In *Graphics interface*, pages 162–167, 2018.

[98] Greg Tamargo. Wiibuddy. https://gregmakesgames.com/wiibuddy/, 2021. Accessed: June 24, 2021.

[99] George Marsaglia et al. Xorshift rngs. *Journal of Statistical Software*, 8(14):1–6, 2003.

[100] Sebastiaan Mathôt and Stefan Van der Stigchel. New light on the mind's eye: The pupillary light response as active vision. *Current directions in psychological science*, 24(5):374–378, 2015.

A STAI & TIS Scores per Participant

	STAI		TIS		
Participant	**State**	**Trait**	**Fear**	**Immobility**	**Total**
P01	23	24	10	4	14
P02	28	37	4	4	8
P03	29	34	11	10	21
P04	24	37	14	20	34
P05	41	45	5	14	19
P06	27	25	7	8	15
P07	32	29	11	21	32
P08	29	28	3	4	7
P09	38	40	8	20	28
P10	25	40	3	3	6
P11	50	26	5	14	19
P12	38	52	10	18	28
P13	26	46	14	6	20
P14	41	48	13	23	36
P15	22	26	7	15	22
P16	36	42	6	29	35
P17	50	58	9	30	39
P18	27	30	4	4	8
P19	34	42	10	16	26
Combined	32.63	37.32	8.11	13.84	21.95

TABLE 10: The STAI and TIS scores per participant.

B Vocal Fear-Intensity Scores per Block and Participant

		Block one		Block two		Block three		Block four	
MarticiMant	**Fear-stimulus**	**Fear**	**Neutral**	**Fear**	**Neutral**	**Fear**	**Neutral**	**Fear**	**Neutral**
P01	Remy	4	0	4	0	4	0	4	2
P02	Josh	1	0	1	0	1	0	2	1
P03	Remy	3	3	2	3	1	2	2	2
P04	Josh	4	2	4	1	4	0	3	0
P05	Remy	2	1	0	1	0	0	1	2
P06	Josh	3	3	2	1	2	2	3	4
P07	Remy	2	0	0	0	1	0	1	1
P08	Josh	1	0	0	0	0	0	0	0
P09	Remy	4	0	4	0	5	0	4	0
P10	Josh	0	0	0	0	0	0	0	0
P11	Remy	4	4	5	5	5	2	5	1
P12	Josh	3	2.5	4	4	3	2	4	2
P13	Remy	3	1	1	1	2	1	1	1
P14	Josh	1	0	0	0	0	0	0	0
P15	Remy	2	2.5	2	0.5	0	0	1.5	0
P16	Josh	3	0	2	0	3	0	1	0
P17	Remy	3	1	2	0	2	0	2	1
P18	Josh	1	0	0	0	0	0	0	0
P19	Remy	5	3	5	2	4	3	3	3
Combined		2.58	1.21	2.00	0.97	1.95	0.63	1.97	1.05

TABLE 11: The fear intensity scores given to each of the stimuli after each block per participant.

C Centers of Balance per Participant During Block One

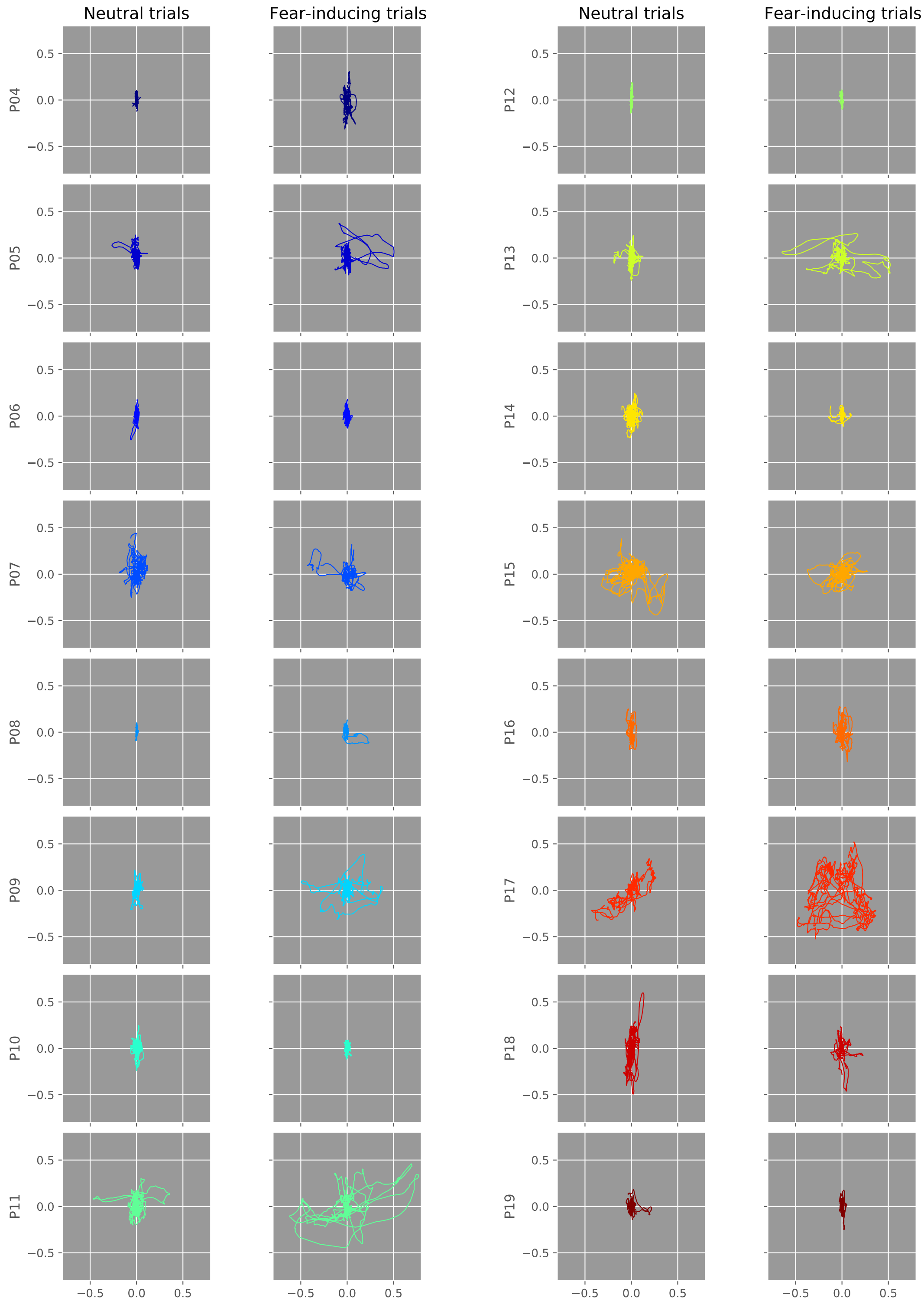


FIGURE 35: Plots of the center of balance paths over time for each participant during block one. These plots give an overview of the general movement behavior of each participant.

D Body Sway Measurements per Participant During Block One

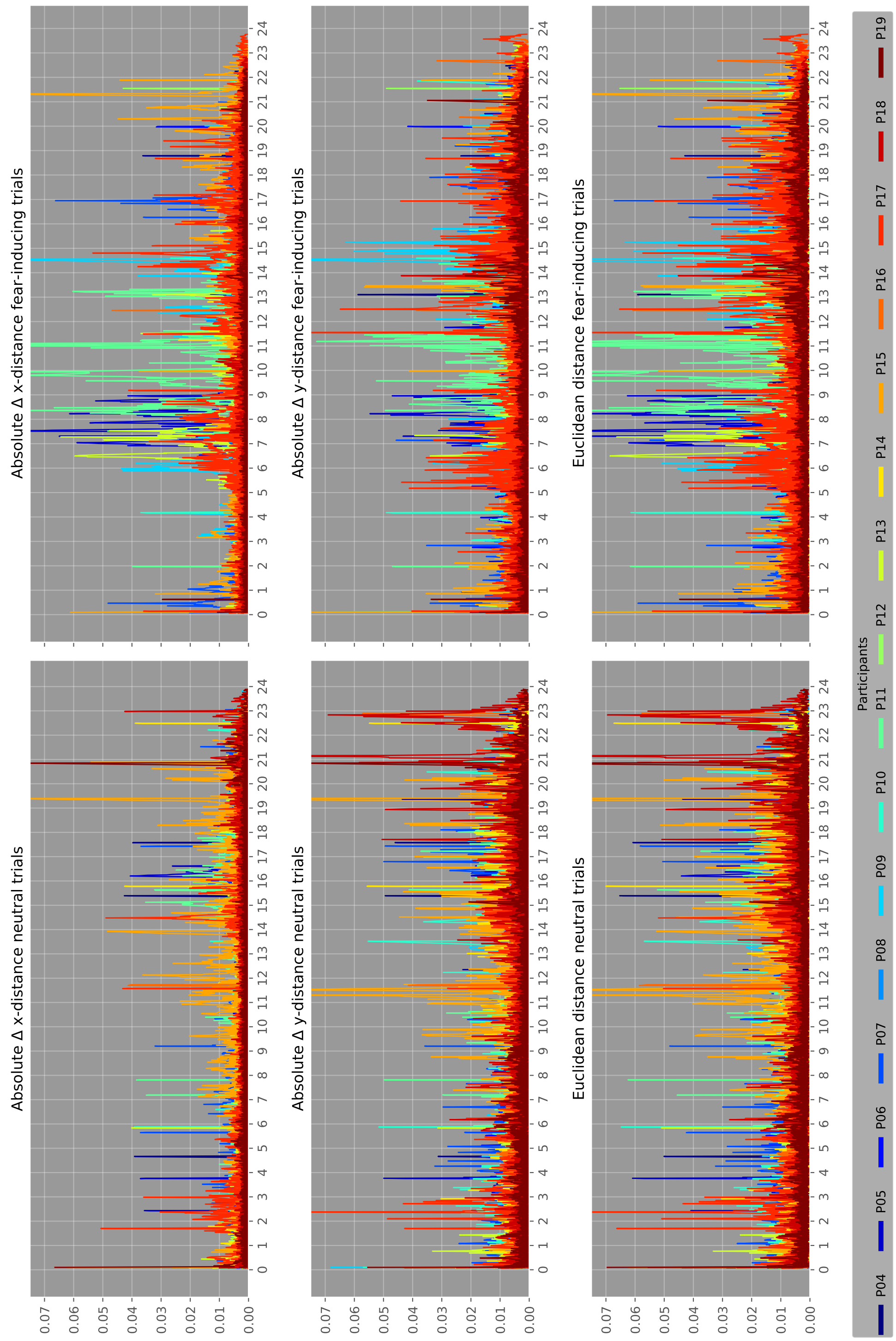


FIGURE 36: Plots of the body sway measurements over time per participant. These plots give an overview of the movement behavior during the trials. Since there were some large peaks for some of the participants, the y-axes have been limited to 0.07.

E Gaze Targets per Participant

Below, the gaze targets are visualized per block, per sub-interval, and per participant. These plots give an overview of the time spent looking at various objects of importance. The category *Nothing* stands for the time in which the participant either blinks or the gaze tracker has lost track of the eyes. Appendix E.1 visualizes the gaze targets of block one, Appendix E.2 visualizes the gaze targets of block two, Appendix E.3 visualizes the gaze targets of block three, and Appendix E.4 visualizes the gaze targets of block four.

E.1 Gaze Targets During Block One

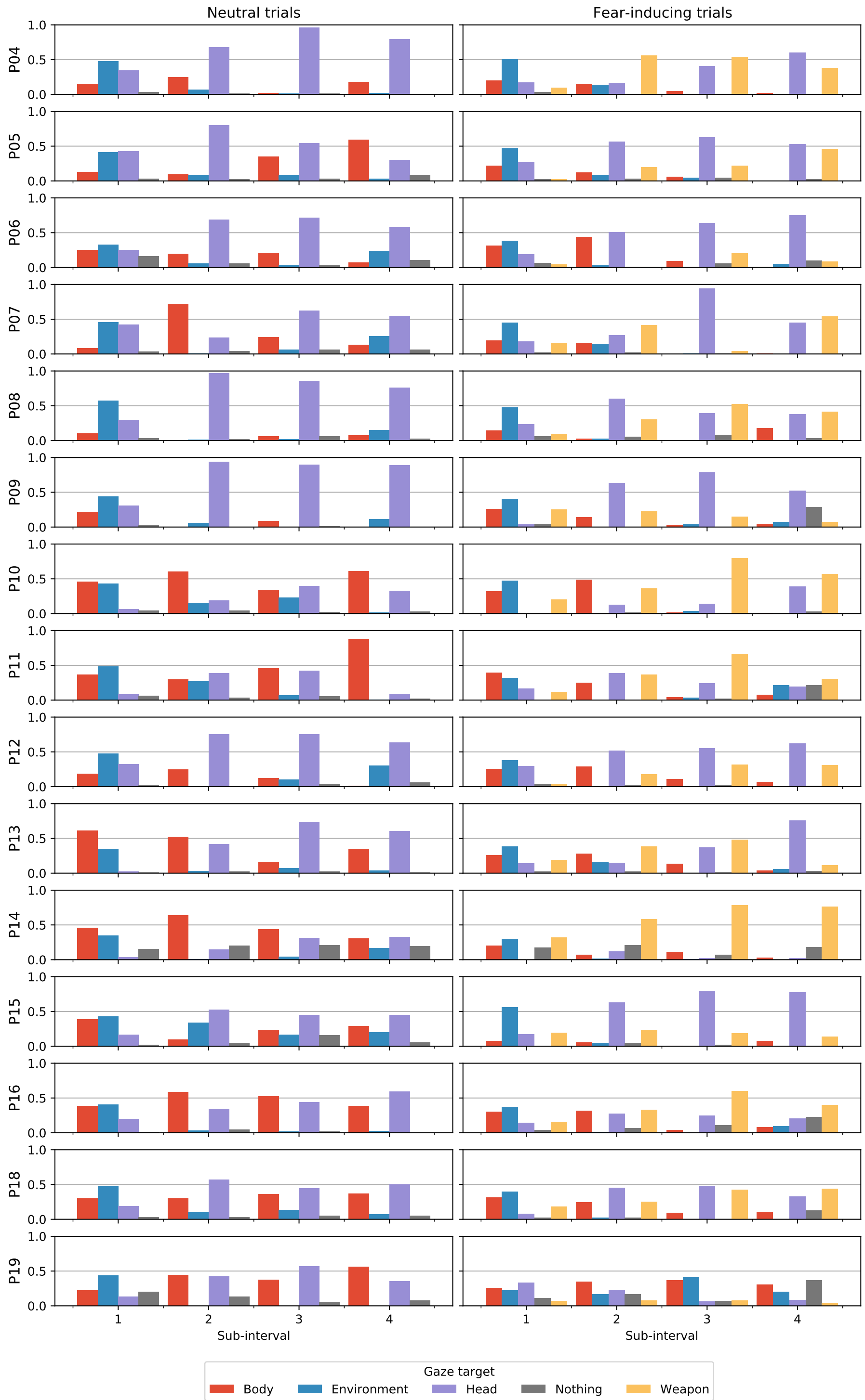


FIGURE 37: Plots of the gaze targets per sub-interval and participant during block one.

E.2 Gaze Targets During Block Two

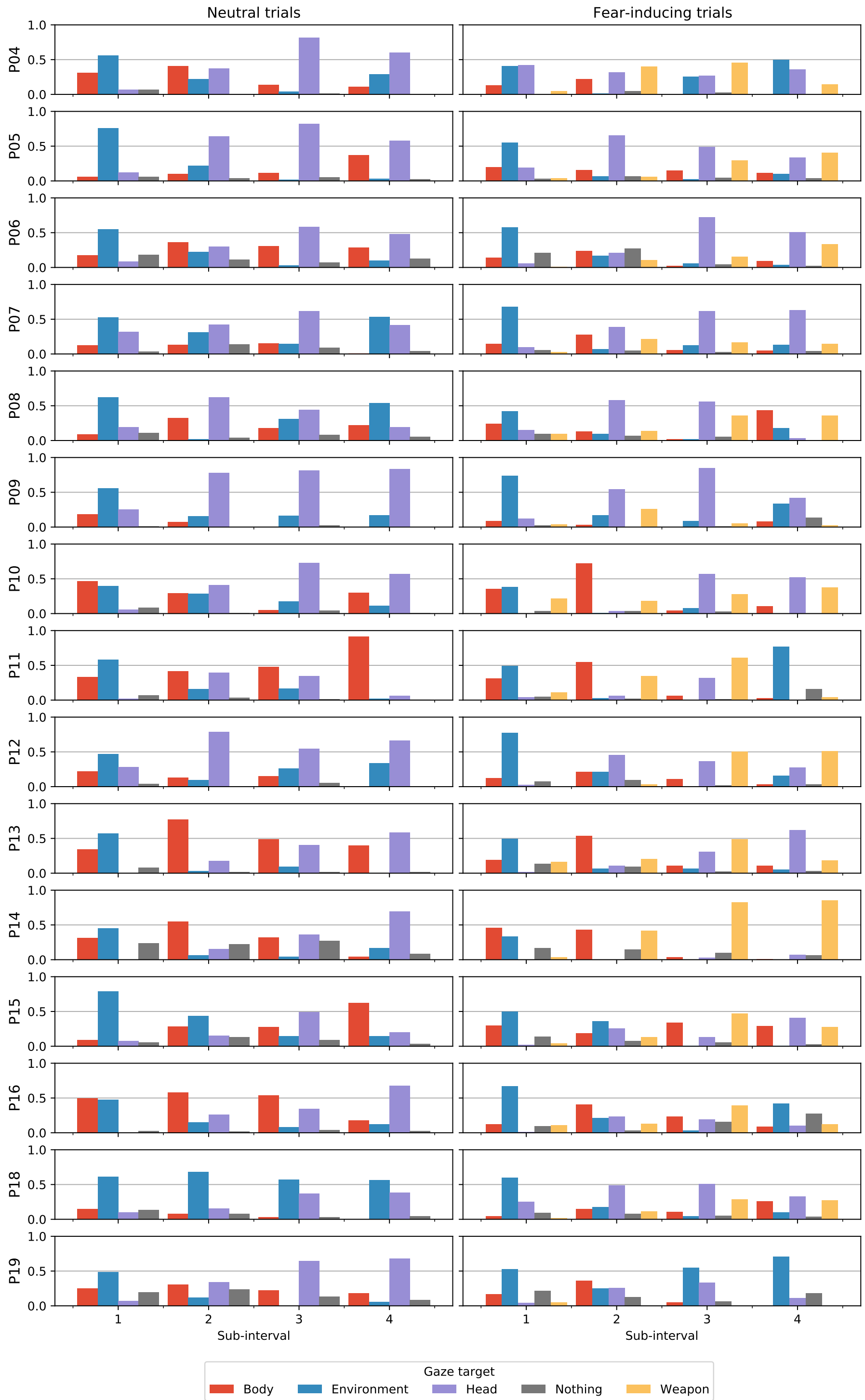


FIGURE 38: Plots of the gaze targets per sub-interval and participant during block two.

E.3 Gaze Targets During Block Three

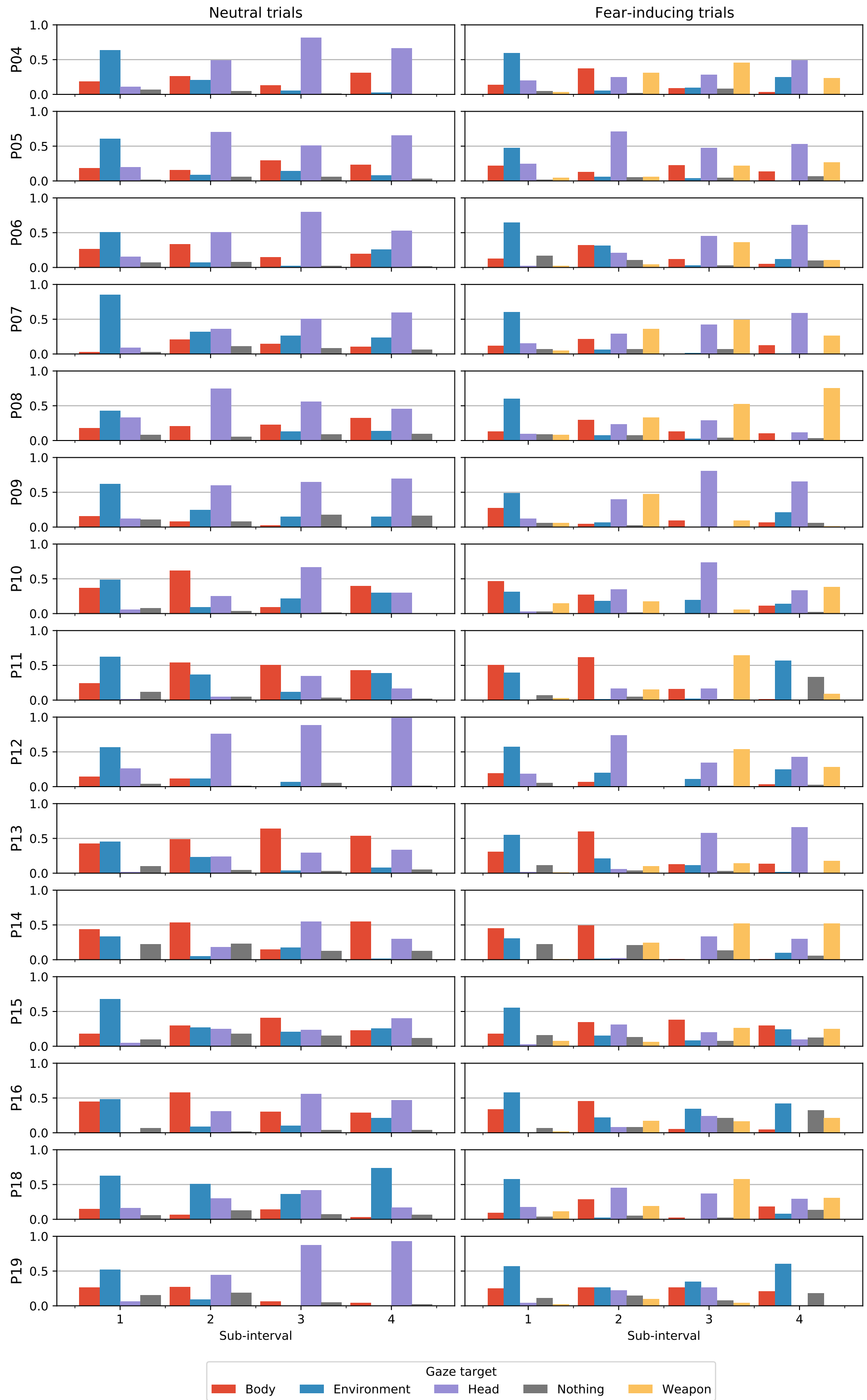


FIGURE 39: Plots of the gaze targets per sub-interval and participant during block three.

E.4 Gaze Targets During Block Four

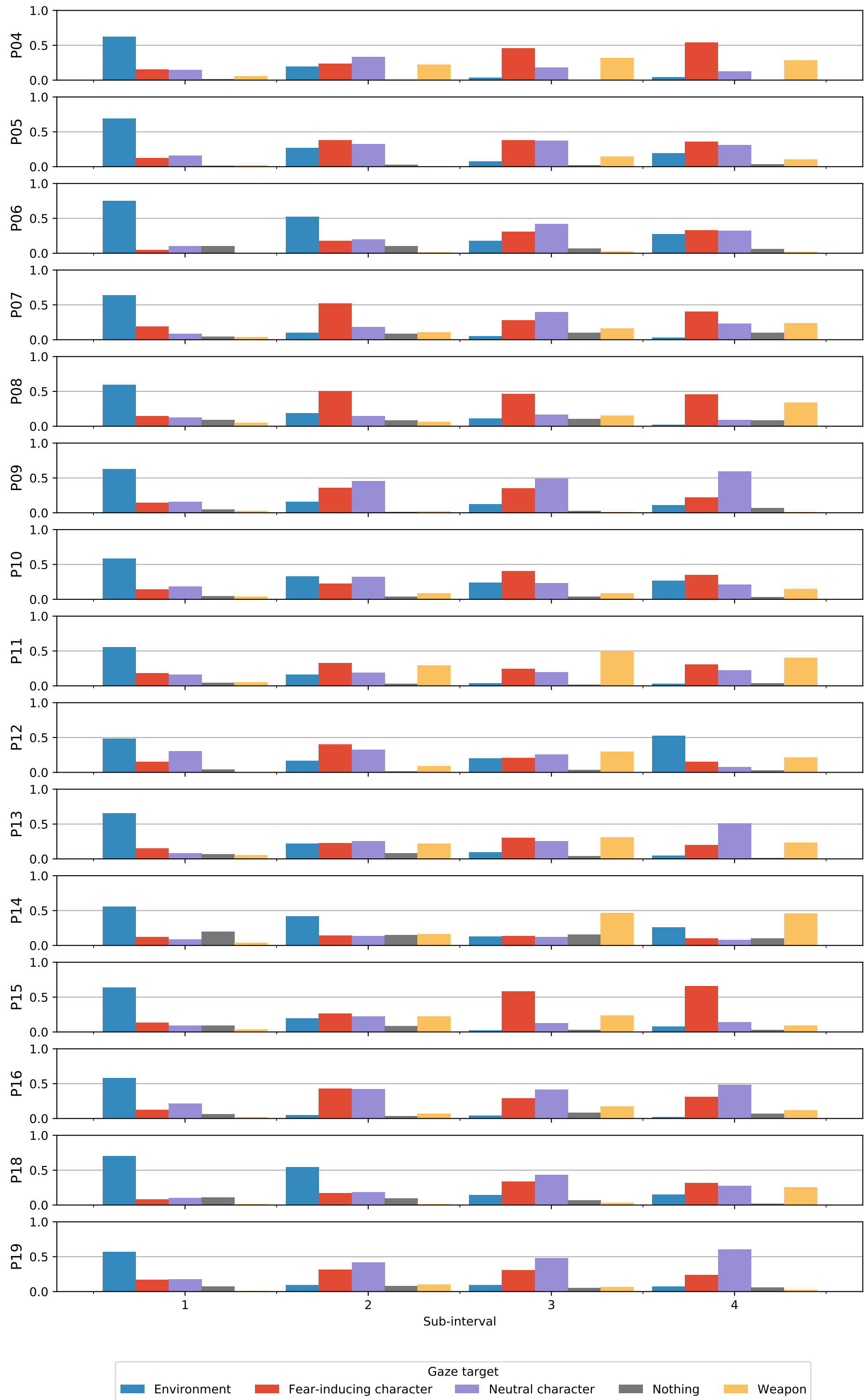


FIGURE 40: Plots of the gaze targets per sub-interval and participant during block four. Here, the body parts are grouped per character. Both types of stimuli were presented simultaneously during this block.