Multisensory Integration in Virtual Reality: Effects of Passive Haptic Stimulation

Master Thesis

submitted in fulfillment of the requirements for the degree
Master of Science (M.Sc.)
in the master's program "Mind and Brain"

Humboldt-Universität zu Berlin
Berlin School of Mind and Brain

Handed in by: Benjamin Dupré
Date of birth: 26.04.1986
Address: Hoppestraße 16, 13409, Berlin

1. Supervisor: Dr. Michael Gaebler
2. Supervisor: Professor Dr. Arno Villringer

Berlin, January 27, 2024

1. Introduction

Interoception intricately weaves together the everyday physiological and cognitive mechanisms responsible for perceiving, comprehending, and integrating internal bodily cues. This continuous mapping of our dynamic internal state, operating across both conscious and unconscious levels, distinguishes interoception from exteroception (perception of the environment) and proprioception (awareness of body position). This often-overlooked aspect illuminates the profound impact of fluctuating internal bodily signals on brain functions. These signals not only relay information about bodily conditions but also significantly influence our perception and interaction with the environment, essentially shaping our reality (Galvez-Pol, Munar, and Kilner 2022).

Immersive Virtual Reality (IVR) operates within this realm by harnessing visual and proprioceptive signals, replacing physical world information with computer-mediated input. It envelops individuals in a world altered by technology, highlighting how VR has the potential to induce a sense of presence within virtual environments. This notion of presence signifies the degree to which individuals realistically respond to these virtual environments, spanning from psychophysical responses to emotional and behavioral reactions (Sanchez-Vives and Slater 2005).

In the realm of psychophysics, studies traditionally explore the correlation between the physical world and its sensory representations. These experiments often involve very simple experimental setups that allow for precise measurement of a single stimulus (intensity, length, etc.). Also, they involve a limited number of highly trained subjects contributing extensive observations (Kingdom 2012; Waskom, Okazawa, and Kiani 2019). Conversely, investigations into higher-level emotional and behavioral aspects typically recruit larger, less specialized cohorts, focusing on population-level analyses (Waskom, Okazawa, and Kiani 2019).

Despite the 40-year evolution of IVR, the majority of research has emulated physical reality, assessing skill transference and similar aspects. However, IVR holds great potential for multisensory research and to look into interoceptive, exteroceptive, and proprioceptive signal integration. Yet, it has been primarily viewed as a medium for simulating existing experiences rather than exploring its relative non-realism and poverty of its perceptual elements. There is an opportunity in VR, in the fact that we can with high precision manipulate sensory information while leaving the rest unchanged. We can build experimental setups to better understand how all signals used by our brain and bodies integrate together to form what we call reality (Vasser and Aru 2020; Gelder, Kätsyri, and Borst 2018).

This thesis proposes an experimental VR setup using a multisensory integration study as reference and platform for exploring this uncharted territory. To be able to use IVR and its "lack of realism" we need to have very present the results in physical world. The subsequent sections delve into the specific findings categorized into psychophysical and cognitive studies, which build up for to the reference study we intend to use as contrast. Later this thisis stress the relevance of using IVR as a new Conitive Psychophics. Which will surely bring ecological validity and broader generalizability beyond experimental settings (Nastase, Goldstein, and Hasson 2020).

Perception & Psychophysics

Psychophysical research explores the relationship between physical phenomena and their perceptual effects. These experiments typically focus on a single perceptual modality, employing high-precision measurements. Within the Interoception literature, a subset of psychophysical studies centers on the cardiac cycle, specifically the events during each heartbeat (systole and diastole). These studies present perceptual stimuli synchronized with a specific cardiac phase (phase-lock) and assess changes in reaction time (RT) or detection task performance.

This body of research indicates that interoceptive signals significantly influence our perception of external signals in touch, vision, and auditory cues. In the somatosensory domain, detecting near-threshold stimuli with higher accuracy occurs during the diastolic phase compared to systole (Al, Iliopoulos, Forschack, et al. 2020; Al, Iliopoulos, Nikulin, et al. 2021; Grund et al. 2022; Motyka et al. 2019). For vision and auditory stimuli, diastole improves accuracy and reaction time compared to systole (Saltafossi et al. 2023). However, real-life stimuli rarely occur in a single modality. Do interoceptive signals impact multisensory integration?

My reference study aims to answer this question by replicating some results from psychophysical experiments, contributing to the validation of using immersive virtual reality (IVR) and passive haptic stimuli. Saltafossi et al. (2023) investigates the integration of multiple simultaneous sensorial modalities (interoceptive and exteroceptive). Forty healthy participants performed a detection task involving unimodal (Auditory, Visual, Tactile) and bimodal (Audio-Tactile, Audio-Visual, Visuo-Tactile) stimuli. These stimuli were presented either 250 ms after the R-peak of the electrocardiogram (systole) or 500 ms after (diastole). The study found a general impact of cardiac cycle phases on detecting both single and combined stimuli, with reaction times being faster during diastole.

Importantly, I employ and replicate the well-known Race Model Inequality (RMI) and response times (RT) to measure multisensory integration. Redundant signal effects in sensory modality testing refer to individuals responding faster to redundant sensory signals than unisensory signals. One proposed explanation is the Race Model Inequality (RMI), suggesting that evidence for each signal accumulates separately in parallel decision units (e.g., one for audition and one for vision), and the first unit to reach its threshold triggers a response (Innes and Otto 2019; Miller 1982).

As measured in my reference study, the diastolic cardiac phase enhances the integration of sensory signals. Audio-Tactile and Visuo-Tactile stimuli show higher integration during diastole compared to systole, unlike Audio-Visual stimuli. This observation suggests a potential specificity in the influence of the cardiac phase on multisensory integration, particularly in stimuli involving somatosensory inputs (e.g., tactile).

All studies so far involve the passive presentation of stimuli and phase-locked conditions. Nevertheless, in our daily lives, stimuli come to us in a constant stream through multiple modalities. Would a more naturalistic stimuli interaction still reveal a modulation role of interoceptive signals? Can VR help us anwer this further?

Interocpetion & Cognition

The second group of studies delves into higher cognitive functions. For instance, one study aims to measure if there's a difference in how we memorize a word presented during a specific cardiac phase. Words were shown within a limited attentional timeframe synchronized with different cardiac phases. This sought to investigate whether natural baroreceptor activity affects word detection and subsequent memory. The study reveals that recalling words presented during systole is lower compared to those presented at diastole. This memory decrease during systole is more pronounced for words identified with low confidence and heightened among individuals with lower interoceptive sensitivity, measured through a heartbeat counting task (Garfinkel et al. 2013).

Another two studies aim to understand if the way in which we actively sample the world is similarly modulated by interoceptive signals. The first study focuses on exploring the role of the heartbeat in active information sampling, investigating whether humans unconsciously arrange their environment to encounter pertinent signals during preferred cardiac phases. In the visual memory experiment's encoding phase, participants navigated through a series of emotional pictures, aiming to memorize them for a subsequent recognition test. Through self-paced key presses, they initiated the display of brief (100 ms) images. The study's findings unveil fluctuations in self-triggered picture onsets throughout the cardiac cycle, notably heightened during cardiac systole, yet without impacting memory performance. This leads the study to conclude that active information gathering incorporates signals related to the heartbeat (Kunzendorf et al. 2019). A similar study confirms these findings. It was achieved by presenting participants with arrays to compare in size. They measured participants' eye movements, heart rate, and response times. The authors similarly found a significant coupling of saccades, subsequent fixations, and blinks with the cardiac cycle. They observed that more eye movement occurs during systolic phases while more fixation happens during diastolic phases, thus demonstrating an active perceptual role (Galvez-Pol, McConnell, and Kilner 2018).

So far we have observed interoceptive signals effects on exteroceptive perception. We have also seen these signals enhance multisensory integration, particularly in stimuli involving somatosensory (e.g., tactile) inputs. Additionally, when observing these interoceptive signals, we have noted an increase in cognitive behavioral outcomes, such as word recognition, and promotion of active information sampling behavior. But how do all of these separate facts integrate together?

Inmersive Virtual Relity as the new tool for psychophysical investigations of cognition

Several previously cited papers propose a theoretical account including to some extent the previously point findings, mostly indicating an interoceptive predictive framework. This framework suggests that repetitive bodily signals, such as respiration cycle, could be predicted and suppressed to avert conscious perception. This inadvertently suppresses external stimuli (Al, Iliopoulos, Nikulin, et al. 2021; Saltafossi et al. 2023; Allen et al. 2022).

One paper presents a comprehensive computational interoceptive predictive framework (Allen et al. 2022). This paper advances the field by employing a Markov Decision Process. The agents deduce their pursued policy (relaxation or arousal) to create the most effective mapping. Operationally defining policies involves transitioning between interoceptive states. For instance, in a relaxed state, the probability shifts among cardiac states, resulting in two phases of diastole and one of systole. In contrast, arousal triggers an immediate shift from the initial diastolic state to systole. Essentially, arousal prompts cardiac acceleration and prolongs the time spent in systole on average. Its premise assumes that precise visual information is available only during specific phases of the cardiac cycle, contingent upon one's arousal state. Among its findings, the paper demonstrates the model's ability to replicate various psychological and physiological phenomena found in the interoceptive inference literature and provides a means to test such a model.

Achieving precise control in investigating the body-brain basis of cognition poses considerable challenges. Cognitive processes are influenced by sensory inputs and manifest in behavioral responses. However, they are distanced from the external variables manipulable by an experimenter and prone to being complex, parallel, and interactive systems. These systems, at best, are only partially understood. Without careful consideration, uncertainty about the effects of an experimental manipulation on the cognitive process of interest persists, accompanied by confounding changes in other processes (Waskom, Okazawa, and Kiani 2019).

To diminish the distance in experiments through cognitive psychophysics, three key points should be adhered to. Firstly, emphasize tasks where expert observers make threshold-level judgments about experimental stimuli enabling precise control. Secondly, adopt a conceptual orientation focused on quantification as a primary goal of experimental design. Finally, place formal computational models centrally in the analysis and interpretation of behavioral and body-brain data (Waskom, Okazawa, and Kiani 2019).

During a virtual reality (VR) experience, an interaction occurs between sensory systems and higher cognition. Leveraging the limitations of the VR world, rather than its richness, provides a unique opportunity to explore the design and mechanisms of brain systems underlying phenomena such as multisensory integration and their intricate characteristics. For instance, a somatosensory signal designed for VR can systematically differ from a real-life somatosensory stimulus. Rather than striving for heightened realism, researchers can create various 'impoverished' versions of real somatosensory signals by systematically manipulating their attributes. Investigating these manipulations is akin to developing and testing a set of hypotheses to unravel the properties of the underlying functional design of multisensory integration and its interoceptive and exteroceptive foundations (Gelder, Kätsyri, and Borst 2018).

In this thesis, the initial phase of this approach involves replicating previous studies and aligning the experimental conditions more closely with real-world situations. This enhances the ecological validity and furthers our comprehension of how body-brain interactions manifest in human psychology (Schmuckler 2001).

The Thesis

We propose a novel experimental setup using Immersive Virtual Reality (IVR), acknowledged for its effectiveness in studying cognitive processes within controlled yet complex scenarios. Traditionally relying on visual displays and head-controller movement tracking, our study introduces a VR head-mounted display with Electrocardiogram (ECG) monitoring and haptic gloves. However, integrating these devices poses challenges, both practical and technical, as well as in terms of alignment with existing literature (Klotzsche et al. 2023).

The primary objective is to validate the manipulation in the context of multisensory integration, with a secondary goal to replicate the RMI findings from Saltafossi et al. (2023). The former enhances our understanding of using IVR as a tool for psychophysical investigations of cognition, while the latter nest our research within previous findings. To my knowledge, the reference study is the only one using RMI analysis to explore multisensory integration, considering interoceptive and exteroceptive signals.

Results indicate participants reported increased immersion with the gloves, although without a significant effect on task performance. Additionally, response times (RT) showed a significant overall difference for the touch condition, validating the setup. The secondary objective revealed overall performance partially adhering to the RSE, notably with the congruent bimodal condition (V = T) significantly faster than the unimodal condition (V). However, no significant differences were found between the incongruent bimodal condition ($V \neq T$) and the single condition (V). Furthermore, examining the cumulative distribution function (CDF) for different RTs replicated the well-known RSE findings, showing faster bimodal RTs than unimodal conditions, with potential violations, specifically in the first 200 ms.

Considerable distinctions exist between the stimuli employed in this study and the referenced study. In this investigation, vision is the sole unimodal signal, while the reference study includes measurements for single modalities in vision, touch, and audio. Additionally, in the reference study, touch is provided as a small electrical charge determined for each participant using the method of limits. In our study, touch is simulated by a vibrotactile glove. For the scope of this thesis inquiry primarily focuses on conditions related to the tactile sense.

By investigating the influence of passive haptic stimuli on response time within a motor-memory task, this study replicates the principal observations outlined in heart cycle multisensory integration literature (Saltafossi et al. 2023), confirming IVR as a means of developing psychophysical investigations of cognition.

2. Methods

Participants

The call for participants targeted healthy, non-smoking German-speaking individuals between 18 and 30 years old. Initially, 23 individuals participated in the study, but ultimately, only 20 participants were included in the final sample. Among the three excluded individuals, two failed to complete all necessary questionnaires, while the third exceeded the age limit. Notably, a greater number of women (15) responded to the call compared to men (5). The average age across the sample hovered around 25 years, with a slight deviation observed in one participant ($\mu = 25.1, \sigma = 6.3$).

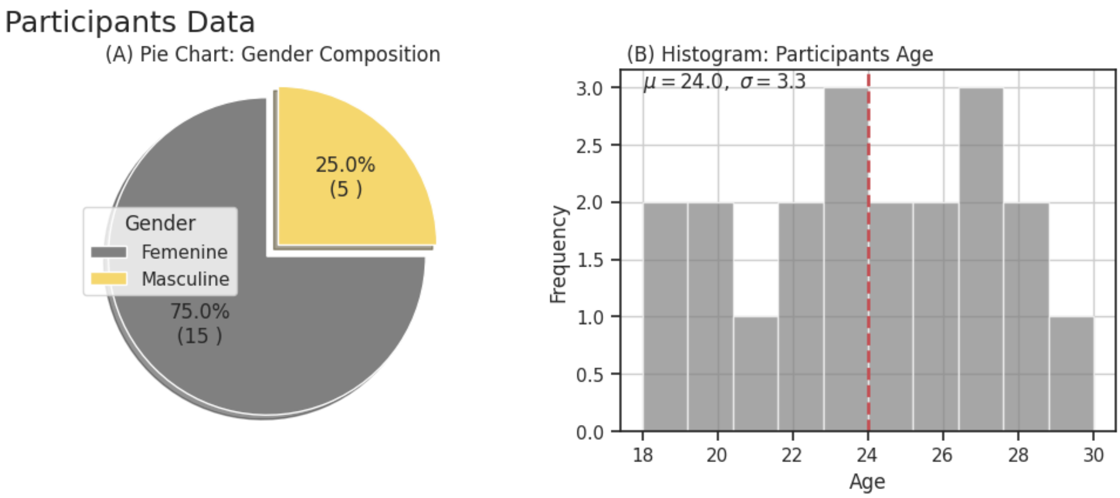

Figure 1: Participants Composition

Materials

Electrocardiogram (ECG):

Heart rate data was collected using an Arduino Uno and a SparkFun Single Lead Heart Rate Monitor - AD8232. The collected data is transferred through a USB 2.0 connection and integrated into the Unity log file at a frequency of 133 Hz. Compared to a clinical ECG, this device entails a serial interface that can send triggers via USB directly to a computer and software (e.g. Unity, Matlab) with minimal delay due to its architecture. Its software and hardware is open-source and publicly available (Möller et al. 2022).

Head Mounted Display & Lighthouses:

The VR setup includes a HTC Vive head-mounted display (HMD) with two lighthouses. The headset specifications include a Dual AMOLED 3.6" diagonal display, with 1080 x 1200 pixels per eye (2160 x 1200 pixels combined), a 90 Hz refresh rate, and a 110-degree field of view. The lighthouses are equipped with SteamVR Tracking, G-sensors, gyroscopes, and proximity sensors. Both the HMD and lighthouses are connected using USB 2.0. For this study, the VR controllers were not used, and instead, hand tracking was performed using the Leap Motion sensor.

Leap Motion Controller:

The Leap Motion Controller has a field of view of 150x120 degrees, with a variable range of roughly 80 cm (arm's length). It weighs 32 grams and is mounted on the HMD. The device features two 640x240 infrared cameras with a frame rate of 120 fps.

Data Gloves:

The data gloves used in the study are equipped with magnetic sensors and connected to Unity using a microUSB connection. These gloves provide haptic feedback through 10 vibrotactile actuators, offering a wide range of tactile sensations with 1,024 levels of intensity. The gloves also incorporate complete finger tracking using six 9-axis Inertial Measurement Units (IMUs). These IMUs enable precise tracking of finger movements, allowing for accurate gesture recognition and enhanced interaction in virtual environments.

Task

Participants, following debriefing on Covid-specific rules, information privacy, and ethical norms, were briefed on the experiment. Subsequently, they consented and completed all subsequent tasks. It's crucial to note that I didn't utilize all tasks from the original study Tactile Stimulation in Virtual Reality (TSVR)" Akbal & Villringer as this thesis pursues a different hypothesis. Nevertheless, considering that all participants completed all questionnaires and tasks, the following description encompasses all tasks for transparency.

Questionnaires: Prior to the IVR experience, participants completed the Edinburgh Handedness Questionnaire to determine their handedness. The PRE-Cybersickness Questionnaire and POST-Cybersickness Questionnaire were administered before and after the IVR task to assess sickness symptoms in participants. Following the IVR experience, participants filled out the Virtual Reality Subjective Evaluation Questionnaire, designed to gather their perceptions of immersion, particularly considering tactile stimuli.

Heartbeat Count Task (HCT): Participants performed a one-minute heartbeat count task before and after the IVR task. Note that this task is not considered within this thesis due to being outside the scope of this secondary study.

After completing the initial questionnaires and HCT, participants moved to another room where the IVR equipment was set up. This included a head-mounted display, data gloves, and an ECG device. Participants received a brief training session before proceeding with the heartbeat count task and the IVR memory-motor task.

IVR Memory-Motor Task: In Figure 2, two panels are depicted. On the left side of the figure (Panel I), we observe the external view of a participant wearing the OVR, ready to commence the IVR Memory-Motor Task. Panel II on the right side of the figure showcases four out of the five steps participants undergo when initiating a trial set. These steps are outlined below:

- **a.** Participants stand in front of a virtual table, allowing ample time to acclimate to the virtual environment, as depicted in Figure 2. When they feel prepared, the session commences as they calibrate by placing their hands on the virtual table.

- **b.** Prior to each trial within every set, a new calibration process begins. Participants place their palms facing up within the shadowed hands, ensuring standardized positioning. Refer to Figure 2 (b) for the calibration setup.

- **c.** Once the calibration is complete, a two-dimensional sketch appears in front of them (Figure 2 (c)), prompting them to memorize the red ball's position. Participants then observe a template on the table, resembling the initial sketch they memorized. Their task is to place the ball swiftly and accurately in the correct location on the template. In the memory-sketch, the relevant position of the ball is denoted by the red circle. During this phase, participants keep their virtual hands open with palms facing up.

- **d.** As soon as the memory-sketch disappears, a red ball appears in either the left or right hand of the participants. Simultaneously, a vibration could start in the glove. The vibration may or may not match the visual location of the red ball. If the vibrating hand matches the visual location of the ball, the condition is congruent ($V = T$). If the vibrating hand does not match the visual location of the ball, then the trial is incongruent ($V \neq T$). If there is no vibration at all, the condition is purely visual (V).

- **e.** After placing the ball on the template, the ball and template disappear. The whole process from (b) to (e) repeats.

The tactile stimuli was an activation of all 10 vibrotactile actuators for 100 ms. The intensity of a vibrotactile pulse used in haptic feedback ranges from 0.0 to 1.0. The set intensity for the program was of 0,2 as coded in Unity Game Engine. All levels had a maximal time of 2 minutes but all participants finished levels before tha cap time by placing the ball before. The parameters, types, and spatial aspects of haptic feedback are configurable, allowing for a versatile setup suitable for psychophysical studies.

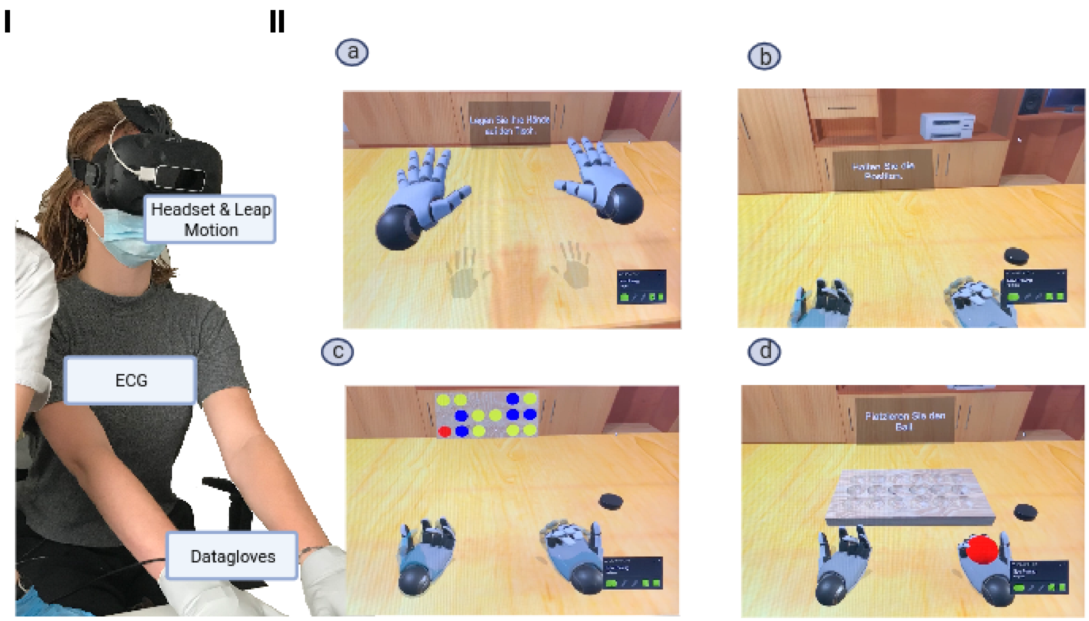

Figure 2: Illustration of the IVR Memory-Motor Task: Panel I show a participant wearing equipment. Panel II shows step (a) represents the acclimation period during initial setup and first step calibration; (b) showcases the first-person perspective of the second calibration phase; (c) displays the 2D sketch for memorizing ball position; (d) presents the appearance of the ball and the 3D template for ball placement, marking the beginning of each trial.

In a sequence of 105 trials, three conditions—congruent ($V = T$), incongruent ($V \neq T$), and visual-only (V)—were presented randomly and rapidly, each occurring 35 times. Each trial commenced upon positioning the ball in a hand, concluding only upon its precise placement on the designated template (refer to Figure 2).

Given that an IVR task inherently involves proprioceptive signals, all conditions in the experiment integrate these signals and are thus multimodal. Consequently, the 'visual only' condition doesn't align with the unimodal classification commonly used in psychophysics literature. Nonetheless, for the sake of easier comparison to this literature in this thesis, I will refer to it as unimodal, and the congruent and incongruent conditions as bimodal.

Mesurements

Immersive Virtual Reality (VR):

Movement data from the data gloves, Leap Motion device, and the HMD was collected. For movement analysis, only the wrist movements tracked by the Leap Motion device were considered, excluding the fingertips' magnetic tracking sensor data. All movements were recorded in a Euclidean coordinate system (X, Y, Z) with the original calibrating point set at (0, 0, 0). This provided a total of nine streaming sources of data (e.g., Headset X, Headset Y, Headset Z, and so on). Notably, rotational data was not included in the analysis.

Additionally, in the game output data, there are flags that signal if a button was pressed, if the ball is placed in the holder, and when the trial started.

Questionnaires:

Both of the described questionnaires are included in the appendix of this thesis for further reference. Questionaire validation is outside the scope of this thesis

(i) **Virtual Reality Subjective Evaluation Questionnaire:** This self-designed questionnaire comprises 26 items aimed at assessing the sense of reality experienced during the VR session. It explores factors like engagement level, hand movement, task difficulty, and other controlling aspects. Participants responded using a Likert scale rating them on a scale from 1 (Does Not Apply) to 7 (Totally Applies). The questions were organized in groups and inverted to confirm participants' responses. Both questionnaires can be found in the Appendix section.

(ii) **PRE/POST-Cybersickness Questionnaire:** This study employs a shortened version of the simulator sickness questionnaire (SSQ) (1993). It utilizes a Likert scale ranging from one to four, featuring labels such as "not present," "somewhat," "clearly," and "very strongly." The questionnaire consists of 16 items, gauging symptoms like "fatigue" and "general discomfort," among others.

Data Analysis

Questionnaire

The analysis of the questionnaire responses aimed to explore the impact of haptic gloves on reported immersion and related perceptions. Initially, the raw questionnaire data collected from 20 respondents, consisting of 27 questions rated on a scale from 1 to 7. Missing values were checked and handled appropriately, thus leaving us with 19. Statistical analysis included the calculation of descriptive statistics such as median (Mdn), mean (μ), and standard deviation (σ) for each question. These statistics were instrumental in understanding the central tendencies and variability in respondents perceptions. A full view of all answers can be found in the Supplements section.

Response Time

All response times from the ball's entry into the scene until its disappearance were measured. Trials where the error button was pressed were excluded from the analysis. Prior to analysis, outliers were corrected by eliminating data points deviating more than 3 times the median absolute deviation (MAD), which is equivalent to 3 standard deviations assuming a normal distribution (Innes and Otto 2019). Responses were not eliminated for being too fast; however, 13% of trials were excluded due to excessive slowness. The final sample comprised 1826 response times (approximately 24 per condition per participant). Subsequent to the removal of outliers, the response times for each condition were transformed into rates (1/RT).

The transformation method entails the inversion of the RT and aligns with prior studies (Innes and Otto 2019). However, this transformation method is not exempt from criticism (Lo and Andrews 2015). This study prioritizes normality in error distribution over other factors such as property scale or interacting effects. This prioritization allowed for the direct application of repeated measures ANOVA within subjects. Yet, when the ANOVA results were ambiguous, a General Linear Mixed-Effect Model (GLMM) was additionally applied, as recommended (Lo and Andrews 2015). By using GLMM, instead of imposing normality and eliminating error deviation, we allow the use of distributions that match the properties of the measured RT (Lo and Andrews 2015). I utilized the *statsmodels* statistical package in Python, specifically the *mixedlm* function, similar to other studies (Swinkels, Veling, and Schie 2021). Although this analysis is not present in the reference study, the utilization of GLMM aims to provide clarity regarding the significance of the conditions and the study's power.

Redundancy Signal Effect and Race Model Inequality

In the cited paper, Race Model approaches suggest that elements from two unimodal stimuli undergo processing in distinct sensory channels, with the fastest one prompting the response (essentially, it "wins the race"). The RMI serves to dismiss the idea that quicker response times (RTs) could be explained by separate processing (i.e., Race Model). It posits that the combined RTs distribution for redundant stimuli never surpasses the total of the RTs distribution for the individual unimodal stimuli. Rejecting this notion indicates interactions across multiple senses.

However, reaching this conclusion in this thesis cannot follow the same methodology as our reference study due to unique characteristics. For example, a constant proprioceptive signal is inherent in the IVR design across all conditions. Additionally, having only one individual unimodal stimulus prevents constructing the complete distribution of reaction times for individual unimodal stimuli since not all modalities are individually measured. Nevertheless, comparisons between the congruent and incongruent conditions with the single modality condition (V) can be made. Furthermore, constructing cumulative distribution functions allows testing the statistical significance of these differences.

Therefore, let's outline the formal definition of our specific case of the race inequality model in equations (1) and (2):

$$F_y(t) \leq F_x(t), \quad t > 0, \quad (1)$$

$$F_z(t) \leq F_x(t), \quad t > 0, \quad (2)$$

Here, F_x represents the cumulative density functions (CDFs) of reaction times (RT) in the individual stimulus conditions visual-only (V), respectively, while F_y and F_z denote the CDF of RT in the redundant-stimulus conditions congruent ($V = T$) and incongruent ($V \neq T$). According to race models, there's a possibility for $F_z(t)$ or $F_y(t)$ to closely approach $F_x(t)$ for small values of t, particularly in scenarios with strongly negative correlations in detection times (2007). However, even under these circumstances, the inequality must still hold true as per the race models.

The process involved generating empirical cumulative density functions (CDFs) for three conditions—Bimodal pairs and Single Signal—and followed the steps outlined in literature (2007). The first steps were followed for every participant and every stimulus condition. Specifically, let G_x be the individual CDF estimate of the visual-only condition, and F_y and F_z denote the redundant CDF for the incongruent and congruent conditions.

Consider a scenario where a set $\{x_1, x_2, \ldots, x_n\}$ of n reaction times (RTs) has been recorded within condition V for a specific participant. Arranging this sample in ascending order, from the smallest value to the largest—$x_1 \leq x_2 \leq \ldots \leq x_n$—one creates a step function. The second step involves using a step function to generate a cumulative frequency polygon. The final step is the estimation of percentiles and aggregation across participants. For a comprehensive breakdown and reference code, consult Ulrich, Miller, and Schröter 2007. Additionally, code for this thesis is shared on GitHub.

To mimic the results from Saltafossi et al. 2023, I conducted a series of t-tests using individual participant data for the F_x, F_y and F_z values at each percentile level. Rather than examining areas under the curve, I focused on the raw values of F_x, F_y and F_z. Furthermore, instead of emphasizing time bins, I opted to discuss percentile levels, given the differing time ranges compared to traditional RMI literature are longer because of answering times in IVR task.

Results

Questionnaires

All participants were right-handed and exhibited a significant increase in reported sickness. A comparison between Pre and Post VR-Cybersickness Questionnaires indicated a rise in the average cybersickness among all experiment participants. Overall results leaned towards symptoms being "not present". We performed a paired-samples t-test to assess the cybersickness levels before (Pre) and after (Post) the Virtual Reality Experience. There was a significant difference in the scores between the Pre (M=1.10, SD=0.12) and Post (M=1.18, SD=0.19) conditions; $t(21) = -2.65$, $p = 0.015$. On average, the symptoms moved from "not present" to "somewhat present".

Virtual Reality Subjective Evaluation Questionnaire

Nineteen respondents answered 27 questions, rating them on a scale from 1 (Does Not Apply) to 7 (Totally Applies). The key facts from the questionnaire are as follows:

In question number 24, the reported score for the question of a perceived increase in immersion due to the haptic gloves was high (Mdn $= 6$, $\mu = 6$, $\sigma = 0.76$). Furthermore, the task was considered enjoyable at least some of the time (Mdn $= 6$, $\mu = 5.4$, $\sigma = 1.04$).

Question 26 revealed that haptic feedback was perceived as either not significantly improving performance or having a neutral effect (Mdn $= 3$, $\mu = 3.6$, $\sigma = 1.49$). Similarly, in question 12, the perception of haptic feedback improving response time was generally rated as neutral to not applicalbe (Mdn $= 3$, $\mu = 3.7$, $\sigma = 1.74$). In question 7, when asked about the impact on results, haptic feedback was perceived as not applicable as well (Mdn $= 6$, $\mu = 6$, $\sigma = 0.76$).

Notably, the assertion that it was very challenging to remember the position of the red ball (Mdn $= 2$, $\mu = 2.9$, $\sigma = 1.45$) was generally disagreed upon. Conversely, in question 8, which asked the inverse question about the ease of remembering the location of the red ball (Mdn $= 5$, $\mu = 5.1$, $\sigma = 1.37$) allign with its counterfactual and it is report as highly easy.

The highest variance was observed in question 13 (Mdn $= 5$, $\mu = 4.2$, $\sigma = 1.79$), indicating that haptic feedback made it easier to place the ball. Similarly, question 18, which inquired about the ease of counting heartbeats at the beginning of the experiment, also showed significant variance (Mdn $= 4$, $\mu = 4$, $\sigma = 1.79$).

Overall, participants reported increased immersion and enjoyment as a result of the added gloves. However, there was no perceived enhancement in response attributed to the presence of the gloves and the IVR ball placing task was considered easy. Moreover, the notably high variance observed in the last two questions may suggest potential confusion among participants regarding these specific queries or significant individual differences in experiences. For a more detailed breakdown, please refer to the supplementary section.

Influence of Touch Stimuli on Response Time

General Performance

Initially, I assessed the mistake rates in ball placement across different conditions to ensure their negligible impact on the experiment. The mistake rates were as follows: in the visual-only condition (V), it was 6.35% ($\pm 0.99\%$, SEM); in the visual incongruent touch condition ($V \neq T$), it was 5.79%). Additionally, a one-way ANOVA (feedback type) indicated no significant effects ($F \leq 0.08$, $p \geq 0.91$). This task is considerably more complex than a simple yes-no detection task. Therefore, to some extent, we anticipate a higher mistake rate. Nevertheless, these rates remain notably low and were not subjected to further analysis.

To investigate whether the experimental manipulations succesfully manipulated RT, I conducted a one-way repeated-measures ANOVA. The test revealed a significant main effect for the stimulus ($F(2, 38) = 3.4$, $p \leq 0.043$, $\eta p^2 = 0.15$). The median response time for the 'Congruent' condition ($V = T$) was the fastest (3236 ms ± 456 ms), followed by the 'Incongruent' ($V \neq T$) condition (3268 ms ± 470), while the slowest was observed in the absence of haptic stimuli 'None' (V) (3284 ms ± 463). Thus, the experimental conditions significantly influenced response times, as illustrated in Figure 3. However, post hoc Tukey HSD tests did not reveal any specific pairwise differences. While an overall difference among conditions was observed, specific pairwise differences were not identified.

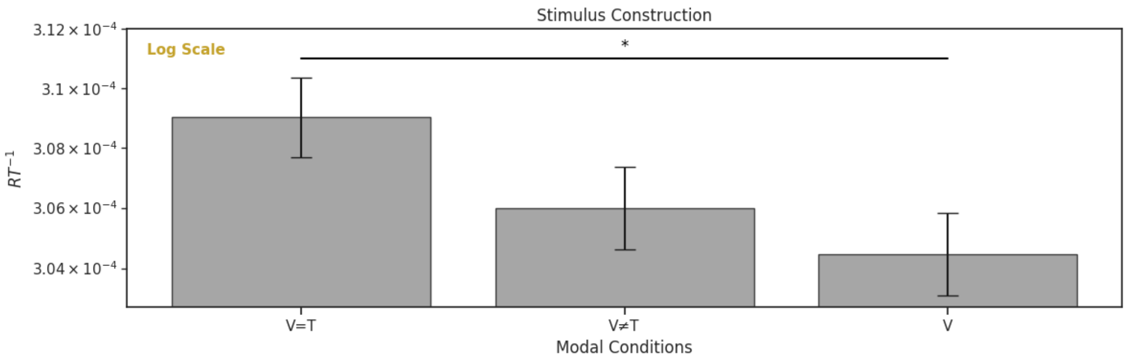

Figure 3: One-way repeated-measures ANOVA using RT^{-1}. The test revealed a significant main effect for the conditions ($F(2, 38) = 3.4$, $p \leq 0.043$, $\eta p^2 = 0.15$)

Given the absence of identified specific pairwise differences in the post hoc analysis, I conducted a more comprehensive investigation using a Generalized Linear Mixed Model (GLMM).

While not employed in the referenced paper, this method offers numerous advantages in multi-level research designs. It addresses the issue of linear relationship between the standard deviation of RT nd mean R, characterized by an increasing spread in residuals for longer predicted RT (Lo and Andrews 2015).

The mixed linear model analysis aimed to assess the impact of feedback types (($V = T$), ($V \neq T$), (V)) on response times. The model included feedback type as a fixed effect and participant as a random effect, with the 'Congruent' feedback type serving as the reference category.

The model coefficient for the 'None' (V) feedback type ($\beta = 45.726, SE = 22.594, p = 0.043$) reached statistical significance at the conventional level ($\alpha = 0.05$) when compared to the 'Congruent' ($V = T$) feedback type. This suggests a significant difference in response times between the 'None' and 'Congruent' feedback types. The coefficient for the 'Incongruent' feedback type ($\beta = 32.975, SE = 22.686, p = 0.146$) did not reach conventional levels of significance.

Thus, the 'Congruent' condition demonstrated significantly faster performance compared to the visual-only condition. This outcome aligns with the expected results according to the RSE. However, the findings present a mixed perspective. Despite this significant contrast, the more stringent post hoc Tukey HSD tests revealed no notable differences between pairwise conditions. Additionally, the GLMM indicated significance solely between the 'Congruent' and visual-only ($V = T$ - V) conditions and not between the 'Congruent' and 'Incongruent' conditions ($V = T$ - $V \neq T$). To understand if this mixed results are due to the irrelevancy of the 'Incongruent' condition ($V \neq T$) or the result of changed behavior over time, CDF analysis will help complement the model.

Race Model Inequeality

In this section, I utilize the Cumulative Distribution Function (CDF) from the Race Model Inequality (RMI) model to evaluate the significance of differences between conditions at different time points (t). The t-test conducted between the three experimental conditions revealed significant differences between the 'Congruent' and visual-only ($V = T$; V) conditions, but no significant differences emerged between the 'Congruent' and 'Incongruent' conditions ($V = T$; $V \neq T$). The results between F_x and F_z (V ; $V = T$) are presented in Table 1, which illustrates the results for all time frames and percentiles. Notably, significant differences emerged between F_x and F_y from the 16th percentile onwards. Interestingly, for this percentile (as depicted in Figure 4), the single modality visual (V) appears to be faster than the bimodal incongruent condition ($V \neq T$), suggesting a potential violation of the Race-Specific Exclusion (RSE) principle. Conversely, from the 40th percentile onwards, both redundant signals consistently demonstrate faster response times compared to the unimodal signal, aligning with our expectations based on the RSE. It's important to note that when considering the distribution of all response times, the highest density is observed between 3100 and 3200 ms, accounting for 33% of all response times.

Discussion

With a completely new IVR set-up, this thesis analyzes the validity of a haptic glove as stimuli in an IVR memory task and compares results to previous multisensory integration literature in the context of RSE and RMI models. By doing so, it takes steps towards bridging the gap between cognitive and psychophysical research, helping IVR become the new Cognitive Psychophysics.

F_x Time Range (ms)	F_x Min Max Diff (ms)	Percentile Estimation	T-value	p-value
(3070, 3621)	551	10	1.19	0.12
(3078, 3626)	548	13	1.31	0.1
(3084, 3648)	564	16	1.88	0.04*
(3086, 3687)	600	19	2.04	0.03*
(3087, 3708)	621	20	2.01	0.03*
(3092, 3748)	656	23	1.97	0.03*
(3094, 3760)	665	26	2.06	0.03*
(3098, 3766)	668	29	2.03	0.03*
(3099, 3768)	668	30	2.02	0.03*
(3104, 3814)	709	33	2.11	0.02*
(3106, 3838)	732	36	2.08	0.03*
(3111, 3846)	735	39	1.83	0.04*
(3114, 3848)	734	40	1.79	0.04*
(3119, 3861)	741	43	1.74	0.05*
(3121, 3907)	786	46	1.55	0.07
(3123, 3969)	846	49	1.36	0.09
(3124, 3978)	855	50	1.33	0.1
(3130, 4171)	1041	60	1.25	0.11
(3141, 4352)	1211	70	1.21	0.12
(3153, 4645)	1492	80	1.47	0.08
(3176, 4948)	1772	90	0.53	0.3

Table 1: T-test results between F_x & and F_z conditions. Response Time Range (ms), Difference RT (ms), Percentile Estimation, T-value, and p-value for each Percentile using the results of t-tests. * indicates significant p-values.

To accomplish this goal, the task required participants to remember a location and immediately place an appearing red ball over this location in a 3D template. The condition is designed so that a passive vibro-tactile stimulus is activated when a red ball visually appears in the hand. The conditions are built as follows: There is a match between the visual location of the ball and the vibration of the glove, then it's called congruent visual-tactile stimuli ($V = T$). When there is a mismatch between the location of the ball and the vibration of the glove, then it's called incongruent visual-tactile stimuli ($V \neq T$). Finally, when there is no vibration in any glove but only IVR visual cues, then the condition is called visual-only stimuli (V). The study employs a within-subject experimental design with the three aforementioned conditions. Additionally, the participants answered a cybersickness questionnaire.

To test if the IVR set-up can mimic existing literature, the steps outlined in my reference studies (Innes and Otto 2019; Saltafossi et al. 2023; Ulrich, Miller, and Schröter 2007) are followed. Using the RSE and RMI models, we can use the cumulative distribution function (CDF) to compare the probability that a response time is equal or less at period time t. By doing so for our three conditions, we can validate the set-up, gain insight into future research, and nest the results in ongoing investigations.

Over a third of all trials were answered within 3100 ms and 3200 ms, making it the most frequent time bin. The first step of analysis involves validating stimulus construction by analyz-

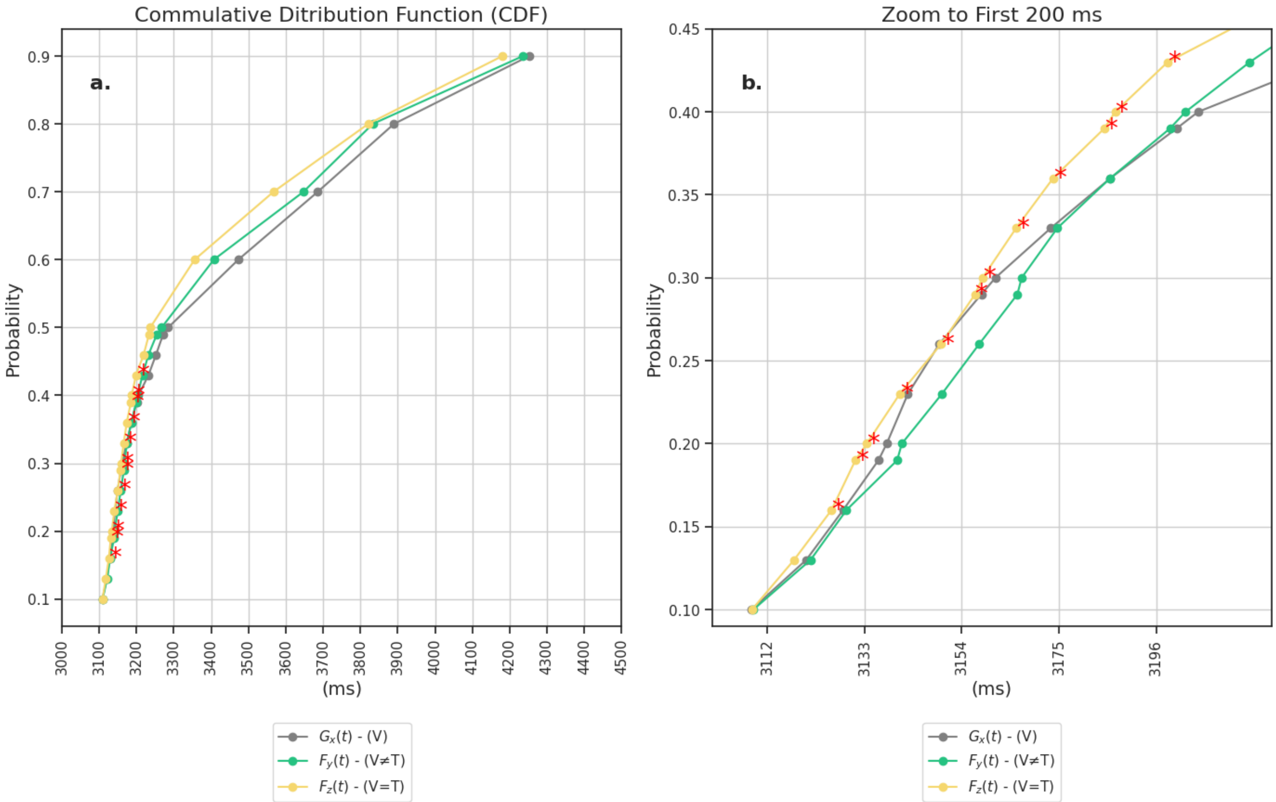

Figure 4: (a.) Displays the estimated percentile points for each of the three functions of interest: G_x, F_y, F_z, considering all participants. (b.) Is a zoom in the first 100 ms. Acording to the RSE, the CDF G_x of the visual-only (V) condition should be under F_y, F_z at all moments. Our data shows that is not the case for the RT in the first 100 ms.

ing and modeling reaction times for unimodal and bimodal stimuli. The results present a mixed picture, indicating an overall significant difference between the conditions. However, while attempting to identify pairwise differences using a Tukey HSD test, no significant distinctions are found. Nevertheless, a GLMM indicates a notable difference between the visual-only (V) and visual-touch congruent ($V = T$) conditions. There are no significant variances observed between the incongruent ($V \neq T$) and visual-only (V) conditions.

The RMI analysis indicates a significant disparity in the cumulative distribution functions (CDF) between conditions when conducting a between-participants t-test for F_x and F_z. This discrepancy occurs specifically within the timeframe of 3100 ms to 3200 ms, rather than before or after. Moreover, no significant differences were observed for F_x and F_y during this 100 ms period or thereafter.

Descriptively, the relationship between the visual-only (V) and congruent condition ($V = T$) remains consistent across the entire range of reaction times (RT), aligning with the Redundant Signals Effect (RSE). However, this pattern doesn't hold for the RT relation between the incongruent condition ($V \neq T$) and the visual-only condition (V), where initially, within the first 100 ms of the trial, the visual-only condition displays faster responses than the incongruent condition ($V \neq T$), potentially breaching the expected inequality. In later response times, the incongruent condition is faster than the unimodal visual-only condition, as anticipated. This again raises the question over a possible violation of the RMI model. Particularly as a bimodal condition ($V \neq T$) demonstrates slower responses compared to a unimodal condition (V).

The question of whether the incongruent condition effectively indicates a violation of the RMI model for the first 100 ms or if it is just a casualty could be addressed through a study with a higher sample of participants. Another study investigating embodiment illusion yielded similar results concerning the irrelevance of stimuli for reaction times (RT). Their hypothesis suggests that the absence of a difference in the Redundant Signals Effect (RSE) between congruent and incongruent conditions could be due to the mere simultaneous occurrence of visual and vibrotactile stimuli. This simultaneous presence might suffice for multisensory integration, regardless of whether these multisensory stimuli are functionally linked at a higher level or not (e.g., touch in the hand holding the ball) (Swinkels, Veling, and Schie 2021). This explanation seems to apply to the responses after the 40th percentile and the overall results, but it doesn't account for the observed first 100 ms, indicating a potential violation of the RMI.

One of this study's limitations comes from the integration of all devices. The latencies between them are not the same, thus creating repeated values for every t in time. This could affect the results in the analysis, especially for psychophysical experimental design. Other considerations need to be held as experiments using IVR will include proprioceptive cues which must be considered when establishing parallels between current and past studies. And finally, each trial in our IVR study spans a longer time frame compared to traditional psychophysics experiments (with a mean duration of 3400 ms versus the typical 200 ms). This extended duration contributes to increased variance and skewness in reaction times (RT).

A new design should pay attention to the latencies between all interfaces (e.g., ECG, Gloves, HMD) and integrate them accordingly, considering that the mean differences between conditions might be between 30 and 50 ms. The experimental design should also measure RT between the trial start and the initiation of movement so the mean RT and variance shorten, thus avoinding the skewness problem and increasing the significance of the results. As well as increasing the sample of future research, involving, for instance, 60 participants.

It would be interesting for future research with an increased sample size to repeat the IVR task and test if the RMI violations have statistical significance or not. Also, if the heart cycle has an effect on RMI violations during multisensory integration.

References

Al, Esra, Fivos Iliopoulos, Norman Forschack, et al. (2020). “Heart–brain interactions shape somatosensory perception and evoked potentials”. In: *Proceedings of the National Academy of Sciences* 117.19, pp. 10575–10584. DOI: 10.1073/pnas.1915629117. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.1915629117. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1915629117.

Al, Esra, Fivos Iliopoulos, Vadim V. Nikulin, et al. (2021). “Heartbeat and somatosensory perception”. In: *NeuroImage* 238, p. 118247. ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2021.118247. URL: https://www.sciencedirect.com/science/article/pii/S1053811921005243.

Allen, Micah et al. (Sept. 2022). “In the Body’s Eye: The computational anatomy of interoceptive inference”. In: *PLOS Computational Biology* 18.9, pp. 1–27. DOI: 10.1371/journal.pcbi.1010490. URL: https://doi.org/10.1371/journal.pcbi.1010490.

Galvez-Pol, Alejandro, Ruth McConnell, and James M Kilner (2018). “Active sampling in visual search is coupled to the cardiac cycle”. In: *Cognition* 196. URL: https://api.semanticscholar.org/CorpusID:91565050.

Galvez-Pol, Alejandro, Enric Munar, and James M Kilner (Oct. 2022). “The influence of interoceptive signals on the processing of external sensory stimuli”. In: *The Routledge International Handbook of Neuroaesthetics*. London: Routledge, pp. 89–102.

Garfinkel, Sarah N et al. (Mar. 2013). “What the heart forgets: Cardiac timing influences memory for words and is modulated by metacognition and interoceptive sensitivity”. en. In: *Psychophysiology* 50.6, pp. 505–512.

Gelder, Beatrice de, Jari Kätsyri, and Aline W. de Borst (2018). “Virtual reality and the new psychophysics.” In: *British journal of psychology* 109 3, pp. 421–426. URL: https://api.semanticscholar.org/CorpusID:44107926.

Grund, Martin et al. (2022). “Respiration, Heartbeat, and Conscious Tactile Perception”. In: *Journal of Neuroscience* 42.4, pp. 643–656. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.0592-21.2021. eprint: https://www.jneurosci.org/content/42/4/643.full.pdf. URL: https://www.jneurosci.org/content/42/4/643.

Healy, Alice F. and Robert W. Proctor (2003). “Handbook of Psychology, volume 4, Experimental Psychology”. In: 4, pp. 147–176.

Innes, Bobby R. and Thomas U. Otto (2019). “A comparative analysis of response times shows that multisensory benefits and interactions are not equivalent”. In: *Scientific Reports* 9. URL: https://api.semanticscholar.org/CorpusID:67862549.

Kennedy, Robert S et al. (1993). “Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness”. In: *The International Journal of Aviation Psychology* 3.3, pp. 203–220. DOI: 10.1207/s15327108ijap0303_3. URL: https://doi.org/10.1207/s15327108ijap0303_3.

Kingdom, F.A.A. (2012). “Psychophysics”. In: *Encyclopedia of Human Behavior (Second Edition)*. Ed. by V.S. Ramachandran. Second Edition. San Diego: Academic Press, pp. 234–239. ISBN: 978-0-08-096180-4. DOI: https://doi.org/10.1016/B978-0-12-375000-6.00296-2. URL: https://www.sciencedirect.com/science/article/pii/B9780123750006002962.

Klotzsche, Felix et al. (2023). “Visual short-term memory related EEG components in a virtual reality setup”. In: *bioRxiv*. DOI: 10.1101/2023.01.23.525140. eprint: https://www.biorxiv.org/content/early/2023/01/23/2023.01.23.525140.full.pdf. URL: https://www.biorxiv.org/content/early/2023/01/23/2023.01.23.525140.

Krakauer, John W. et al. (2017). “Neuroscience Needs Behavior: Correcting a Reductionist Bias”. In: *Neuron* 93.3, pp. 480–490. ISSN: 0896-6273. DOI: https://doi.org/10.1016/j.neuron.2016.12.041. URL: https://www.sciencedirect.com/science/article/pii/S0896627316310406.

Kunzendorf, Stella et al. (Jan. 2019). “Active information sampling varies across the cardiac cycle”. en. In: *Psychophysiology* 56.5, e13322.

Lo, Steson and Sally Andrews (Aug. 2015). “To transform or not to transform: using generalized linear mixed models to analyse reaction time data”. en. In: *Front Psychol* 6, p. 1171.

Mahoney, Jeannette R and Joe Verghese (May 2019). “Using the Race Model Inequality to Quantify Behavioral Multisensory Integration Effects”. en. In: *J Vis Exp* 147.

Miller, Jeff (1982). “Divided attention: Evidence for coactivation with redundant signals”. In: *Cognitive Psychology* 14.2, pp. 247–279. ISSN: 0010-0285. DOI: https://doi.org/10.1016/0010-0285(82)90010-X. URL: https://www.sciencedirect.com/science/article/pii/001002858290010X.

Möller, T. et al. (2022). “An Arduino Based Heartbeat Detection Device (ArdMob-ECG) for Real-Time ECG Analysis”. In: *2022 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pp. 1–3. DOI: 10.1109/SPMB55497.2022.10014819.

Motyka, Paweł et al. (2019). "Interactions between cardiac activity and conscious somatosensory perception". In: *Psychophysiology* 56.10, e13424. DOI: https://doi.org/10.1111/psyp.13424. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/psyp.13424. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/psyp.13424.

Nastase, Samuel A., Ariel Goldstein, and Uri Hasson (2020). "Keep it real: rethinking the primacy of experimental control in cognitive neuroscience". In: *NeuroImage* 222, p. 117254. ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2020.117254. URL: https://www.sciencedirect.com/science/article/pii/S1053811920307400.

Proske, Uwe and Simon C. Gandevia (2012). "The proprioceptive senses: their roles in signaling body shape, body position and movement, and muscle force." In: *Physiological reviews* 92 4, pp. 1651–97. URL: https://api.semanticscholar.org/CorpusID:1512089.

Saltafossi, Martina et al. (2023). "The impact of cardiac phases on multisensory integration". In: *Biological Psychology*, p. 108642. ISSN: 0301-0511. DOI: https://doi.org/10.1016/j.biopsycho.2023.108642. URL: https://www.sciencedirect.com/science/article/pii/S0301051123001606.

Sanchez-Vives, Maria V. and Mel Slater (2005). "From presence to consciousness through virtual reality". In: *Nature Reviews Neuroscience* 6, pp. 332–339. URL: https://api.semanticscholar.org/CorpusID:19179383.

Schmuckler, Mark A (2001). "What is ecological validity? A dimensional analysis". In: *Infancy* 2.4, pp. 419–436.

Swinkels, Lieke M. J., Harm Veling, and Hein T. van Schie (2021). "The Redundant Signals Effect and the Full Body Illusion: not Multisensory, but Unisensory Tactile Stimuli Are Affected by the Illusion". In: *Multisensory Research* 34.6, pp. 553–585. DOI: https://doi.org/10.1163/22134808-bja10046. URL: https://brill.com/view/journals/msr/34/6/article-p553_1.xml.

Ulrich, Rolf, Jeff Miller, and Hannes Schröter (May 2007). "Testing the race model inequality: An algorithm and computer programs". In: *Behavior Research Methods* 39.2, pp. 291–302. ISSN: 1554-3528. DOI: 10.3758/BF03193160. URL: https://doi.org/10.3758/BF03193160.

Vasser, Madis and Jaan Aru (2020). "Guidelines for immersive virtual reality in psychological research." In: *Current opinion in psychology* 36, pp. 71–76. URL: https://api.semanticscholar.org/CorpusID:219441124.

Waskom, Michael L., Gouki Okazawa, and Roozbeh Kiani (2019). "Designing and Interpreting Psychophysical Investigations of Cognition". In: *Neuron* 104.1, pp. 100–112. ISSN: 0896-6273. DOI: https://doi.org/10.1016/j.neuron.2019.09.016. URL: https://www.sciencedirect.com/science/article/pii/S0896627319307895.

Additional Material

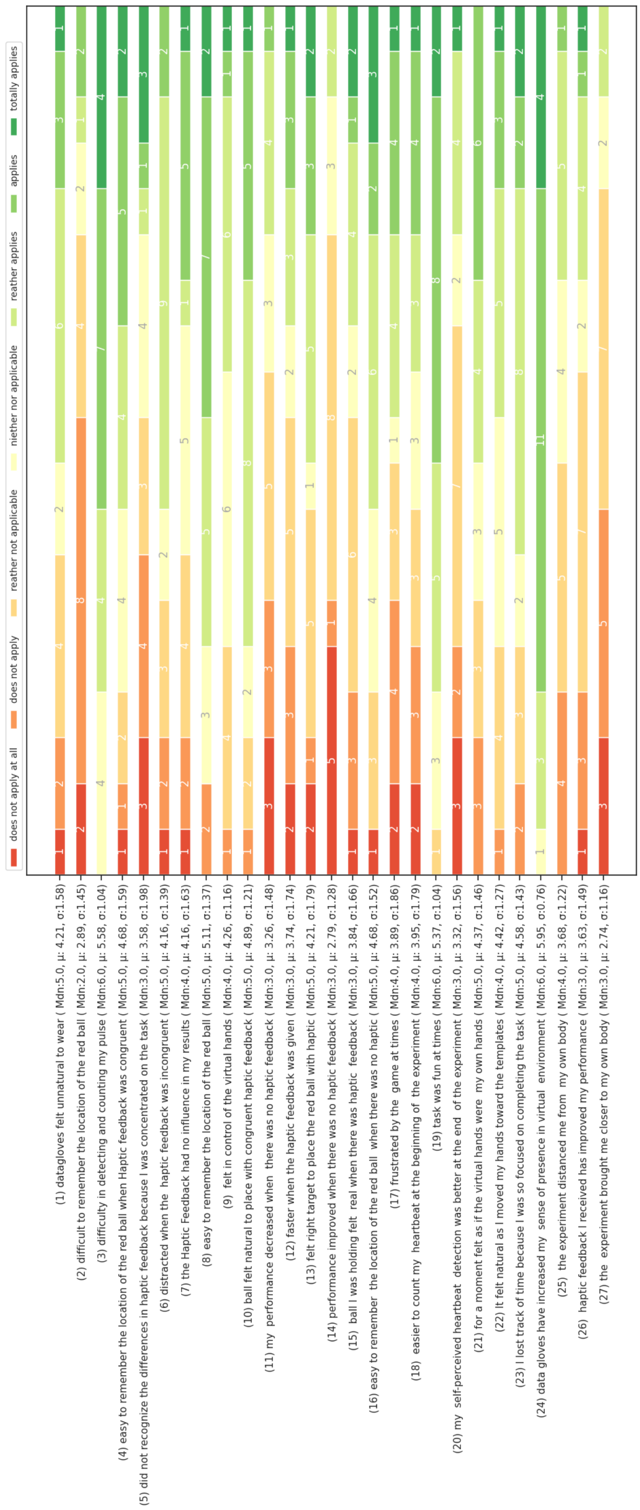

Figure 5: Results: Virtual Reality Subjective Evaluation Questionnaire