Multisensory Integration in Virtual Reality: Effects of Passive Haptic Stimulation

Master Thesis

submitted in fulfilment 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, March 5, 2024

Acknowledgments

I would like to express my sincere gratitude to Dr. Michael Gaebler and Prof. Dr. Arno Villringer for their advice and support in making this thesis possible. Many thanks to Max Hellrigel-Holderbaum for his friendly editing and references, as well as to Dr. Zeynep Akbal and Dr. Tim Julian Möller for generously sharing their expertise and time.

I am deeply grateful for the support provided by the Max-Planck Institute, specifically from Dr. Zeynep Akbal and Prof. Dr. Arno Villringer (NRO-228), and Study-DB (0218807).

1. Introduction

Immersive Virtual Reality evokes psychophysical reactions

Immersive Virtual Reality (IVR) is a computer-generated environment that simulates real-world interactions through electronic equipment. In IVR, visual, auditory, and tactile signals can be replaced with computer-mediated inputs, enveloping individuals in a virtual environment. In other words, it gives the user a sense of physical presence within a digitally constructed world. Presence, in this context, refers to the sensation of being within the virtual environment rather than the physical space one occupies. It is precisely this 'suspension of reality' that makes immersive experiences able to elicit a range of responses from individuals. They span between emotional, behavioural, and psychophysical reactions (e.g. subjective perceptual response to a stimulus with described physical characteristics) (Sanchez-Vives and Slater 2005). These responses, akin to those elicited by traditional media such as theatre or films, often exhibit heightened intensity within IVR. Thus, making it a compelling tool for investigating human behaviour, cognition, and perception.

One can view IVR as a continuation of a long psychophysical tradition that attempts to interfere with our perception to clarify its underlying mechanism. For example, before similarly lenses or prisms have been used to invert visual perception of orientation to study the visual properties of the cortex (Gelder, Kätsyri, and Borst 2018). Nevertheless, if we are to use IVR as a tool to do psychophysical perception studies, the question of its validity and how it compares to real-life arises. This question, "whether responses in IVR are real or not" is a philosophical question that remains beyond the scope of this thesis. Nonetheless, recent empirical research provides insights into this question, by comparing psychophysical and other relevant markers in three different experimental setup responses elicited by IVR compared to real-life experiences or conventional experimental setups.

More importantly, the fact that IVR can simply parameterize stimuli to precisely manipulate available information and computational models to jointly quantify behaviour and neural responses, could play a fundamental role in how we elicit psychophysical reactions for the study of cognitive functions(Waskom, Okazawa, and Kiani 2019).

How real are psychophysical responses in IVR? And how comparable is IVR to what we experience in real-life or in existing experimental setups? A recent study took on these questions by comparing three experimental set-ups: One in physical reality, another in virtual reality and a last one as a typical experimental set-up displayed on a computer screen. The authors looked at various markers like electrophysiological, physiological, and subjective responses (e.g. ECG, ECG, Surveys). What they found was that our brain and body seem to react similarly in both real-life and virtual situations. Specifically, alpha- and theta-band oscillations in line with heart rate variability, indexing vigilance, and anxiety were barely indistinguishable between VR and real-life, while they differed significantly from computer screen display setup. Sensory processing, as reflected by beta-band oscillations, exhibits a different pattern for all conditions, indicating further room for improving VR on a haptic level (Schöne et al. 2023). It is precisely this haptic limitation that this thesis uses as an experimental manipulation.

Not all signals generated in IVR feel entirely authentic to the user, especially when considering haptic feedback as a substitute for real-life touch. But to hold this lack of realness as a criticism to discard results found in lack of it would naively assume that VR is useful for research as it delivers every time more realistic displays of sensory information. Yet, our understanding of perception reveals that it often transcends mere physical stimuli. A realistic representation arises from the intricate interplay of multisensory integration and cognitive processes (Gelder, Kätsyri, and Borst 2018).

However, IVR's primary experimental advantage lies in its precision in manipulating sensory input and its capacity to induce presence, despite users being aware of the virtual environment's unreality. This state, known as the 'suspension of disbelief,' allows for a coexistence of awareness of VR's artificiality with a belief in its experiential reality (Slater 2009). Consequently, our reactions in IVR align with behavioural and psychophysiological responses in real-life indicating that similar cognitive and emotional mechanisms are activated in both real-life and virtual scenarios (Vasser and Aru 2020; Gelder, Kätsyri, and Borst 2018).

Exploring Multisensory Integration in Immersive Virtual Reality: Leveraging Haptic Stimulation

In recent years, the advancement of virtual reality (VR) technology has predominantly centred around visual experiences. However, with the emergence of immersive virtual reality (IVR), which incorporates computer-mediated inputs to include proprioceptive signals alongside visual cues, the display environment has evolved into a crossmodal realm. Crossmodal stimuli, which simultaneously engage two or more sensory modalities, often lead to multisensory integration. This phenomenon refers to the synergistic interaction and fusion of sensory information (Stein and Stanford 2008).

Assessing multisensory integration typically involves evaluating the effectiveness of crossmodal stimulus combinations compared to their individual components in evoking responses. This can manifest as either multisensory enhancement or depression, depending on whether the crossmodal event heightens or diminishes event saliency (Stein and Stanford 2008). For example, studies on bimodal and trimodal stimuli, combining auditory, visual, and tactile cues, have demonstrated faster response times compared to single-modality stimuli (Diederich and Colonius 2004). On average, bimodal stimuli elicited responses 30 ms faster than the fastest single modality response (Diederich and Colonius 2004). In general, crossmodal stimuli are further enhanced or depressed depending on the location and timing of their occurrence.

When a ball is dropped into our left hand, we naturally expect to feel the ball in the same hand and not in the right. This seamless integration of sensory inputs from different modalities is a fundamental aspect of multisensory processing. Central to this process is the ability to effectively sort and link information originating from the same event. Contextual or semantic congruence plays a pivotal role in this synthesis. Research has consistently shown that stimuli conveying congruent information across different sensory modalities lead to faster and more accurate task performance compared to incongruent stimuli. This phenomenon underscores the importance of contextual consistency in facilitating efficient multisensory integration (Laurienti et al. 2003).

Theories such as the Race Model and the Coactivation Model have been proposed to explain the observed speedup in response to redundant signals. The Race Model suggests that redundancy gains stem from "statistical facilitation," while the Coactivation Model posits that signals from different channels contribute to a common pool of activation Miller (1982). To test these models, Miller introduced the Race Model Inequality (RMI), a mathematical criterion that, if violated, supports the Coactivation Model. The RMI posits that crossmodal stimuli can elicit faster reaction times than the fastest single stimuli.

The standardization of experimental setups in IVR has become increasingly important (Vasser and Aru 2020; Gelder, Kätsyri, and Borst 2018; Schöne et al. 2023; Wiesing, Fink, and Weidner 2020), positioning IVR as an advantageous platform for crossmodal studies. IVR devices, such as HTC Vive paired with Unreal Engine and SteamVR, offer low latencies and high precision in stimulus presentation and critical time measurements. Furthermore, IVR enables the manipulation of sensory modalities while maintaining a naturalistic experience. By combining IVR with electrophysiological measurement devices, researchers can gain valuable insights into multisensory integration (Vasser and Aru 2020).

Current study and hypothesis

In this study we set up to answer: Can we use IVR and haptic gloves to study multisensory integration? When we perform an RMI analysis, How do our results compare to existing literature? Most importantly I use an IVR-memory task that requires participants to look to a prompt with the correct position and later on place the ball in the correct hold. The haptic gloves are the manipulation and there is a case without haptic feedback, with congruent haptic feedback and with incongruent haptic feedback. I run an ANOVA for the mistake rate and RT differences to validate the condition. To compare the RMI results to the literature I use the cumulative distribution functions for each condition at a given response time.

In summary, IVR holds immense promise in cognitive neuroscience research, particularly in the domain of multisensory integration. Its ability to closely mimic real-life experiences, coupled with precise control over sensory stimuli and comprehensive data collection capabilities, positions IVR as a powerful tool for unravelling the complexities of touch in multisensory integration. This thesis aims to contribute to the validation of IVR as a valuable instrument for studying multisensory integration, with a specific focus on touch as a computationally mediated signal manipulated using haptic gloves.

2. Methods

Participants

The call for participants targeted healthy, non-smoking German-speaking individuals between 18 and 30 years old. They were contacted and recruited via e-mail, using the database of Max-Plank CBS for registered participants. Participants were required to be righthanded and without neurological or psychiatric illnesses (e.g. epilepsy). When taking part in the experiment participants were paid 9 € the hour. The experimental session took 2.5 hours on average. The participant came to the laboratory in October 2020 the study was conducted following the declaration of Helsinki. All participants provided informed written consent.

Initially, 23 individuals answered and 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 of 30. 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 ($\mu = 25.1, \sigma = 6.3$).

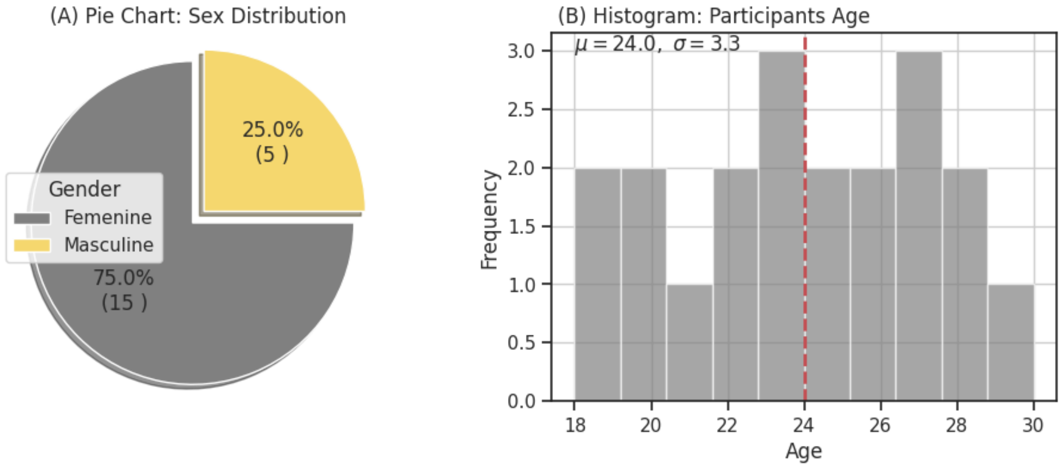

Figure 1: Participant's Distribution

Materials

Electrocardiogram (ECG):

Heart rate data was collected using an Arduino Uno and a SparkFun Single Lead Heart Rate Monitor - AD8232. The data was 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 are open-source and publicly available (Möller et al. 2022).

Head Mounted Display & Lighthouses:

The VR setup included a HTC Vive head-mounted display (HMD) with two lighthouses. The headset specifications included 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:

This device was used to track the position of the hands. 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.

Haptic Data Gloves:

The data gloves used in the study are equipped with magnetic sensors and connected to Unity using a micro USB 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. The datagloves and finger tracking were interfaced from the experimental code using the *UnityDll.Motion* and *C#* NeuroDigital licensed code (NeuroDigital Technologies, 2018).

Task

In the following, I describe all tasks performed by participants. It's important to note that I didn't utilize all tasks from the original study. Nevertheless, considering that all included participants completed all questionnaires and tasks, the following description encompasses all tasks for transparency.

Questionnaires: Before 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. [1]

IVR Memory-Motor Task: After completing the initial questionnaires and HCT, participants moved to another room where the IVR equipment was set up. As mentioned this included a head-mounted display, data gloves, and an ECG device. For the implementation of the experiment, we used the Unity software (v2018.3.11; Unity Technologies, San Francisco, United States) in combination with the SteamVR Unity plugin (Valve Corporation, Bellevue, United States). To synchronize the data streams (i.e., behavioural reports, hand positions, finger positions, ECG), we used custom C# scripts and network-based communication (i.e., timestamps).

The tactile stimuli was an activation of 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 0,2 as coded in Unity Game Engine. All levels had a maximal time of 2 minutes but all participants finished levels before the 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.

The virtual environment simulated a rectangular office with a window. Participants appeared sitting in front of a table. Just over the table, there is an initial prompt indicating them to place their hands over the table.

In Figure 2 are two panels. On the left side of the figure (Panel I), we observe the external view of a participant wearing the head device, ready to commence the IVR Memory-Motor Task.

[1] Note that this task is not considered within this thesis due to being outside the scope of this secondary study.

Panel II on the right side of the figure showcases four out of the five steps participants undergo when initiating a trial set. The 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. Before 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 2b for the calibration setup.

c. Once the calibration is complete, a two-dimensional sketch appears in front of them (Figure 2c), 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 visual condition matches the tactile condition (V = T) and therefore the trail is congruent. If the vibrating hand does not match the visual location of the ball, then the visual condition does not match the tactile condition and the trial is incongruent ($V \neq T$). If there is no vibration at all, the trail 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.

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.

Measures

Immersive Virtual Reality (VR):

Movement data from 3 devices was recorded (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 an Euclidean coordinate system (X, Y, Z) with the original calibrating point set

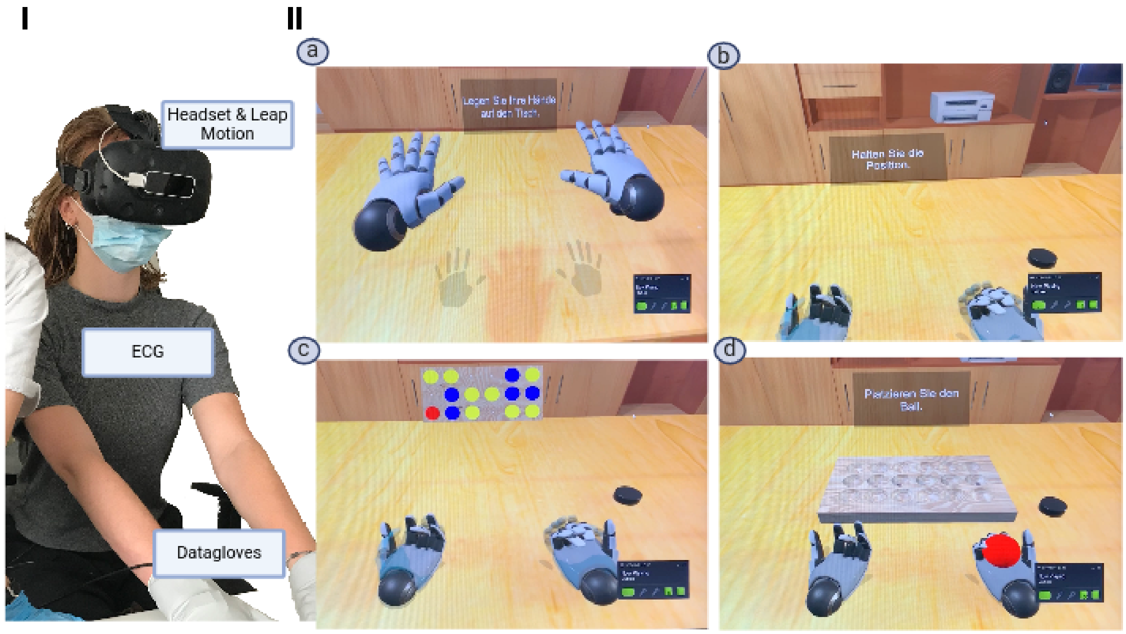

Figure 2: IVR Memory-Motor Task: Panel I shows a participant wearing equipment. Panel II shows the different stages of one trial. (a) participants first view. Here they are asked to place the hands in the virtual table to start the calibration process. 2 seconds after placed the screen changes; (b) the second calibration phase adjusts wrists and fingers. 2 seconds after successfully superposing the hands in the signaled position the screen changes (c) while holding this position, participants are presented with a display of a 2D sketch for memorizing ball position; (d) red ball is introduced and the 3D template for ball placement, marking the beginning of each trial.

at (0, 0, 0). As each device had three coordinates, this provided a total of nine streaming sources of data (e.g., Headset X, Headset Y, Headset Z, and so on). Additionally, rotational data was recorded. Both sources are not included in the analysis since at the moment it does not help to compare with the race inequality model.

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. Trials where the error button was pressed were excluded from the analysis.

Questionnaires:

Both of the described questionnaires are included in the appendix of this thesis for further reference. Questionnaire 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 (Kennedy et al. 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

Virtual Reality Subjective Evaluation

The analysis of the questionnaire responses aimed to explore the impact of haptic gloves on reported immersion and related perceptions. Initially, raw questionnaire data were collected from 20 respondents, consisting of 27 questions rated on a scale from 1 to 7. Some participants did not answer all questions, thus creating specific missing values that were omitted. 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. Before 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). After 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 by 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

The Race Model statement is simple. It suggests that two different unimodal stimuli (e.g., vision, and touch) undergo processing in distinct sensory channels and race to trigger detection. If these two stimuli were presented simultaneously, then we would expect the fastest one to elicit the response (essentially, the one "winning the race"). Therefore, we would be right to expect inequalities (1) and (2) to be held:

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

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

A violation of the inequalities would require that two modalities presented at the same time have a faster response time than one of these modalities presented alone. This would indicate that the two modalities interacted in some way, thus explaining the faster response than a single modality on its own. Accurate knowledge of this is basic for understanding multisensory integration better.

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 (Ulrich, Miller, and Schröter 2007). However, even under these circumstances, the inequality must still hold as per the race models.

It is important to mention here that IVR does not allow the same experimental methodology as traditionally held in race model inequality studies. For example, a constant proprioceptive signal is inherent in the IVR design across all conditions, thus making absolute single modality impossible. Additionally, having only one single unimodal stimulus (vision) available prevents us from constructing the complete set of cases (single touch, single vision). Nevertheless, I make comparisons between the congruent and incongruent conditions with the single modality condition (V) based on the fact that they all have proprioceptive modality included, thus cancelling the effect out. Furthermore, constructing cumulative distribution functions allows for testing the statistical significance of these differences.

The process involved generating empirical cumulative density functions (CDFs) for three conditions—Bimodal pairs and Single Signal—and followed the steps outlined in literature (Ulrich, Miller, and Schröter 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, the code for this thesis is shared on GitHub.

To mimic the results from our reference paper 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

Cibersickness Questionnaire

We performed a paired-sample 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 Virtual Reality Subjective Evaluation Questionnaire are as follows:

The gloves seem to enrich the reported expirience as in question number 24 of the questionnaire, the reported score on the perceived immersion due to the haptic gloves was "Applies" (Mdn $= 6$, $\mu = 6$, $\sigma = 0.76$). Furthermore, question 19 over whether the task was considered enjoyable at least some of the time was also answered as "Applies" (Mdn $= 6$, $\mu = 5.4$, $\sigma = 1.04$).

Although then enriched expirience does not translate into percieved improved performance. 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" (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 $= 6$, $\mu = 6$, $\sigma = 0.76$).

Overall the memory task is percieved as not difficult. 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, question 8, which asked the inverse question about the ease of remembering the location of the red ball (Mdn $= 5$, $\mu = 5.1$, $\sigma = 1.37$) aligns with its counterfactual and it is reported as highly easy.

I suspect that questions about the easiness of the action are not clearly made. Relative to all other questions 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. Furthermore, the significant variation noted in participants' responses to the last two questions may suggest that some individuals are unsure about how easy it is to perform the specific task mentioned. 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 (balls placed in a different position than the one qued in the prompter) 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 indicated no significant effects ($F \leq 0.08$, $p \geq 0.91$) between all three experimental conditions (congruent, incongruent and visual-only). 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 successfully affected the 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.

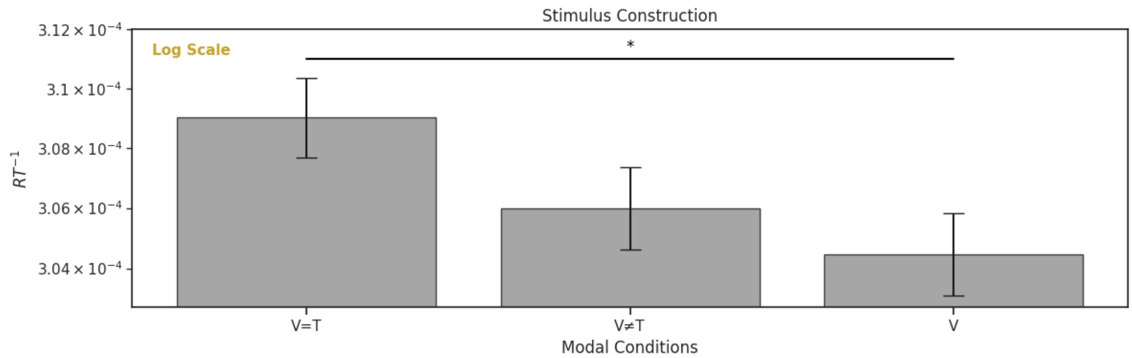

Figure 3: One-way repeated-measures ANOVA using RT^{-1}. The test revealed a significant main effect for the conditions ($F(2, 38) = 3.4$, $p \leq 0.043$, $\eta p^2 = 0.15$)

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

Although this specific analysis method was not employed in the referenced paper (Saltafossi et al. 2023), it offers numerous advantages for multilevel research designs. It addresses, for example, the issue of the linear relationship between the standard deviation of RT and mean RT, characterized by an increasing spread in residuals for longer predicted RT (Lo and Andrews 2015).

The model coefficient for the 'None' (V) condition ($\beta = 45.726, SE = 22.594, p = 0.043$) reached statistical significance at the conventional level ($\alpha = 0.05$) when compared to the 'Congruent' ($V = T$) condition. This suggests a significant difference in response times between the 'None' and 'Congruent' conditions. The coefficient for the 'Incongruent' condition ($\beta = 32.975, SE = 22.686, p = 0.146$) 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' ($V = T$) and visual-only (V) conditions and not between the 'Congruent' ($V = T$) and 'Incongruent' ($V \neq T$) conditions. We should consider that median time differences between conditions range between 30 to 50 milliseconds. Now that we have validated the condition. To gain a deeper understanding of the results and nest them in the literature the next section of RMI and CDF analysis will aim to achieve this.

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 RMI 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

I have carefully analyzed the validity of a haptic glove as a stimuli in an novel IVR memory task. Analysed and compared RT to the results to previous research on multisensory integration in the context of RSE and RMI models. This two results contribute to bridge the gap between psychophysical and cognitive studies and validate the use of IVR as a tool for developing more advanced cognitive psychophysics.

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. As mentioned in the task section, the condition is designed so that a passive vibrotactile stimulus is activated when a red ball visually appears in the hand. The conditions are built as follows: There is a match between

Percentile Estimation	F_z Time Range (ms)	F_z Min Max Diff (ms)	F_z T-value	F_z p-value	F_y Time Range (ms)	F_y Min Max Diff (ms)	F_y T-value	F_y p-value
10	(3070, 3621)	551	1.19	0.12	(3121, 3907)	786	1.1	0.14
13	(3078, 3626)	548	1.31	0.1	(3123, 3969)	846	0.79	0.22
16	(3084, 3648)	564	1.88	0.04*	(3130, 4171)	1041	0.92	0.18
19	(3086, 3687)	600	2.04	0.03*	(3141, 4352)	1211	0.96	0.18
20	(3087, 3708)	621	2.01	0.03*	(3153, 4645)	1492	0.96	0.17
23	(3092, 3748)	656	1.97	0.03*	(3176, 4948)	1772	0.99	0.17
26	(3094, 3760)	665	2.06	0.03*	(3202, 5060)	1857	1.04	0.16
29	(3098, 3766)	668	2.03	0.03*	(3221, 5184)	1963	1.33	0.1
30	(3099, 3768)	668	2.02	0.03*	(3241, 5290)	2049	1.34	0.1
33	(3104, 3814)	709	2.11	0.02*	(3264, 5391)	2127	1.02	0.16
36	(3106, 3838)	732	2.08	0.03*	(3287, 5558)	2271	0.95	0.18
39	(3111, 3846)	735	1.83	0.04*	(3311, 5739)	2427	1.13	0.14
40	(3114, 3848)	734	1.79	0.04*	(3336, 5884)	2548	1.19	0.12
43	(3119, 3861)	741	1.74	0.05*	(3361, 6067)	2706	1.16	0.13
46	(3121, 3907)	786	1.55	0.07	(3396, 6271)	2946	0.85	0.2
49	(3123, 3969)	846	1.36	0.09	(3431, 6503)	3181	0.48	0.32
50	(3124, 3978)	855	1.33	0.1	(3466, 6723)	3421	0.29	0.39
60	(3130, 4171)	1041	1.25	0.11	(3499, 6989)	3680	0.91	0.19
70	(3141, 4352)	1211	1.21	0.12	(3530, 7310)	3939	0.4	0.35
80	(3153, 4645)	1492	1.47	0.08	(3560, 7674)	4253	-0.54	0.7
90	(3176, 4948)	1772	0.53	0.3	(3589, 8081)	4632	-0.27	0.61

Table 1: 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. Each row represents a percentile. Each t-test evaluates a significant difference compared to G_x, * indicates significant p-values.Estimations are for F_y and F_z

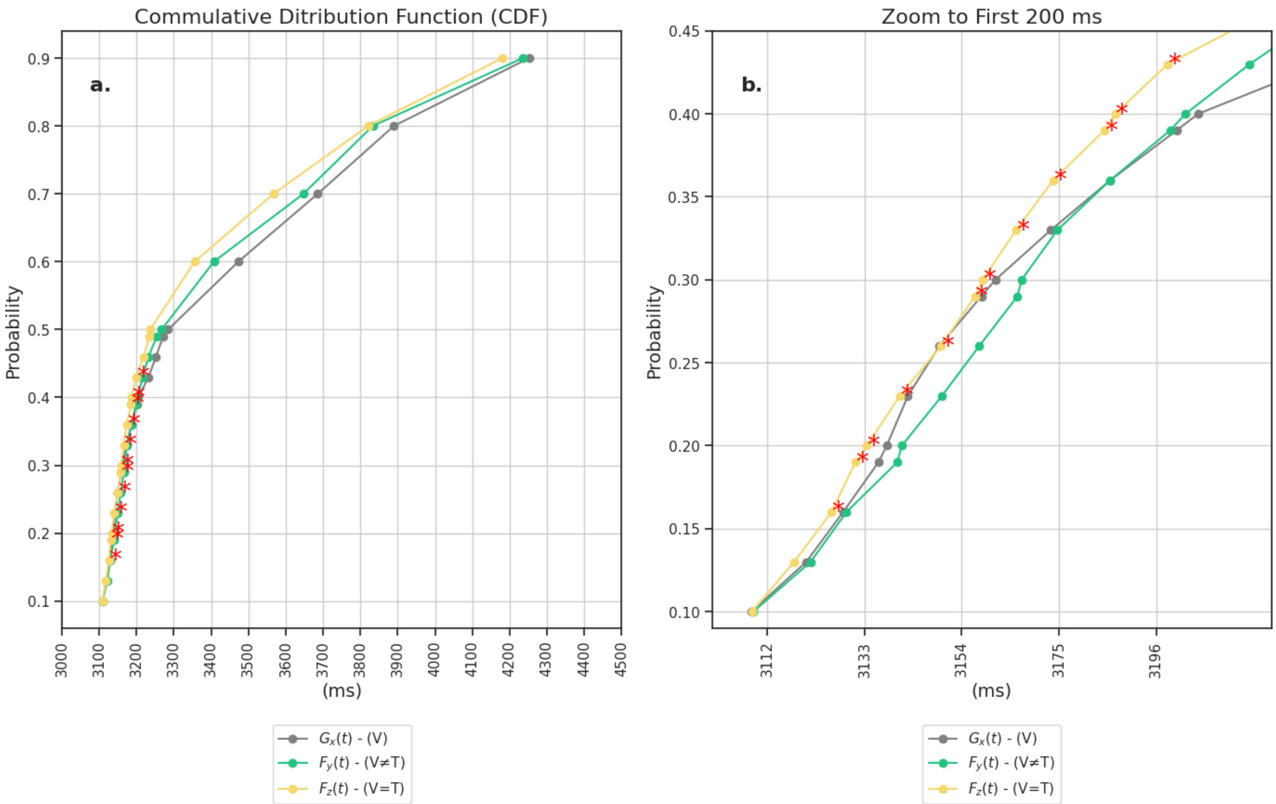

Figure 4: (a.) Displays the estimated percentile points for each of the three functions of interest: G_x, F_y, F_z, considering all participants. (b.) Is a zoom in the first 100 ms of the entire RT range. Acording to the RSE, the CDF G_x of the visual-only (V) condition should be under F_y, F_z at all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

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.

Over a third of the trials were answered within 3100 ms and 3200 ms. To validate stimulus construction, I analyzed and modeled the reaction times for unimodal and bimodal stimuli. The results showed a significant difference between the conditions, but no significant distinctions were found when attempting to identify pairwise differences using a Tukey HSD test. However, a GLMM indicated a notable difference between the visual-only (V) and visual-touch congruent ($V = T$) conditions. No significant variances observed between the incongruent ($V \neq T$) and visual-only (V) conditions.

To compare the IVR set-up with existing literature I followed the steps outlined in my reference study (Saltafossi et al. 2023). 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 against existing psychophysical literature.

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 Race Model inequality (RMI). However, this pattern doesn't hold for the RT relation between the incongruent condition ($V \neq T$) and the visual-only condition (V), where within the first 100 ms of the complete RT range, the single visual condition displays faster responses than the bi-modal 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 of a possible violation of the RMI model. Particularly as a bimodal condition ($V \neq T$) demonstrates slower responses compared to an unimodal condition (V).

I have identified evidence suggesting that the incongruent condition contradicts the predictions of the RMI model within the initial 100 milliseconds of the reaction time (RT) range. However, given the significance lies at the boundary and the effects are relatively minor, I am inclined to seek additional validation to strengthen this interpretation. Another study by Swinkels, Veling, and Schie 2021 investigating embodiment illusion (the reported feeling of seeing one body where is not) yielded similar results concerning the irrelevance of stimuli for reaction times (RT). Their hypothesis suggests that the absence of a difference in the Race Model Inequality (RMI) 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 40th percentile of answers, indicating a potential violation of the RMI.

Some of this study's limitations come from the integration of different devices. The latencies between them are not the same, thus possibly creating data loss or duplicate 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. 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) which needs to be addressed as I did in this thesis. Nonetheless, this makes it harder to compare to other psychophysics experiments.

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 somatosensory conditions might be as small as 20 ms. Thus latency integration issues over 7 ms might hinder the results. The experimental design should also measure RT between the trial start and the initiation of movement so the mean RT and variance shorten, thus avoiding the skewness problem and increasing the significance of the results. Additionally, increasing the sample of future research, involving, for instance, 60 participants to be able to come to more definite conclusions regarding the central hypotheses investigated here (RSE, RMI). A further interesting question to investigate is whether the heart cycle has an effect on RMI violations during multisensory integration.

References

Diederich, Adele and Hans Colonius (2004). “Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time”. In: *Perception & Psychophysics* 66, pp. 1388–1404. URL: https://api.semanticscholar.org/CorpusID:1860837.

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.

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.

Laurienti, Paul J. et al. (2003). “Cross-modal sensory processing in the anterior cingulate and medial prefrontal cortices”. In: *Human Brain Mapping* 19. URL: https://api.semanticscholar.org/CorpusID:24621333.

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.

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.

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.

Schöne, Benjamin et al. (2023). “The reality of virtual reality”. In: *Frontiers in Psychology* 14. URL: https://api.semanticscholar.org/CorpusID:256936423.

Slater, Mel (2009). “Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, pp. 3549–3557. URL: https://api.semanticscholar.org/CorpusID:6521266.

Stein, Barry E. and Terrence R. Stanford (2008). “Multisensory integration: current issues from the perspective of the single neuron”. In: *Nature Reviews Neuroscience* 9, pp. 255–266. URL: https://api.semanticscholar.org/CorpusID:54549705.

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.

Wiesing, Michael, Gereon R. Fink, and Ralph Weidner (Apr. 2020). “Accuracy and precision of stimulus timing and reaction times with Unreal Engine and SteamVR”. In: *PLOS ONE* 15.4, pp. 1–24. DOI: 10.1371/journal.pone.0231152. URL: https://doi.org/10.1371/journal.pone.0231152.

Additional Material

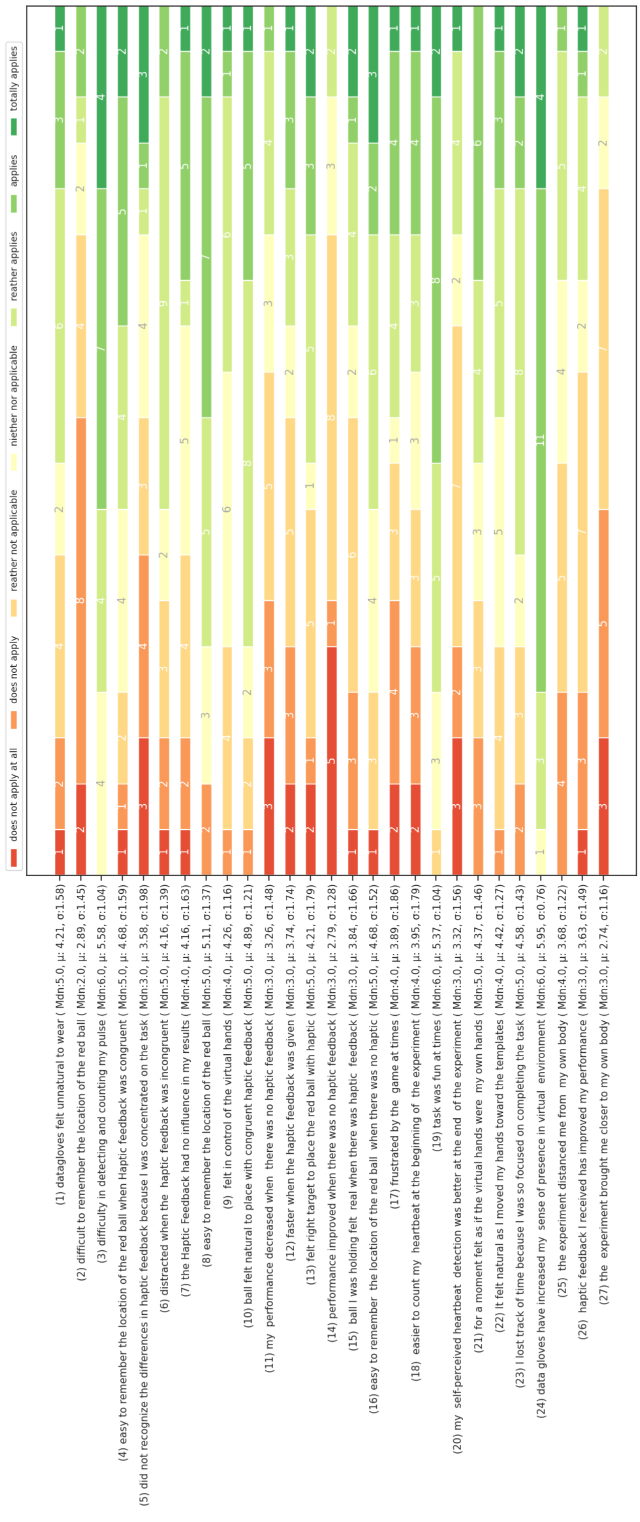

Figure 5: Results: Virtual Reality Subjective Evaluation Questionnaire