BERLIN
SCHOOL OF
MIND AND
BRAIN

HUMBOLDT-UNIVERSITÄT · ZU BERLIN ·

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 7, 2024

Acknowledgments

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' what makes immersive experiences able to elicit a range of responses from individuals. This, responses span between emotional, behavioural, and psychophysical reactions (Sanchez-Vives and Slater 2005). We will understand for psychphysical reactions the subjective perceptual response to a stimulus with described physical characteristics. All of 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.

IVR can be seen as a continuation of a long psychophysical tradition that attempts to interfere with our perception to clarify its underlying mechanism. In the past for example, 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) and IVR would be somewhere along this lines. Nonetheless, if we are to use IVR as a tool to do psychophysical perception studies, many will question its validity and how it compares to real-life. While this issue is in part philosophical and extends beyond the scope of this thesis, recent empirical research offers insights into the comparison of IVR, real-life and computer base experimental setups for psychophysical and other relevant markers.

How comparable is IVR to what we experience in real-life or in current experimental setups? A recent study took on these questions by comparing three experimental set-ups: One in physical reality, another in virtual reality and the last one displayed on a computer screen and keyboard setup. The participants needed to climb up into a firefighter hydraulic platform presented to them via the three above mentioned setups, while 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, heart rate variability, indexed 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). This last result is in line with what we know about IVR. That in general touch mediation or haptics has been overlooked due to methodological challenges. Nevertheless, the sense of touch provides sensory feedback about temperature, texture and rigidity, and has a crucial role in creating a truly immersive virtual environment (Zhang et al. 2023).

As touch mediation in IVR shows, not all computer-mediated signals feel entirely real to users or have an equivalent physiological response to real-life. But to hold this lack of realness as a criticism to IVR lead research would assume that IVR is useful for research thanks to its realness. Yet, our understanding of perception reveals that it often transcends mere physical stimuli. Among other factors, realistic representation emerges from the interplay of multisensory integration and cognitive processes (Gelder, Kätsyri, and Borst 2018). IVR can facilitate the exploration and

manipulation of both these processes.

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). Furthermore, the fact that IVR can quantify stimuli precisely and display an environment rich enough to manipulate cognitive functions allows us to join physical data, subjective perceptual responses and behavioural outcomes in one setup. All this input could allow for computational models that jointly quantify behaviour and neural responses., which could play a fundamental role in closing the gaps between psychophysics and cognitive sciences. (Waskom, Okazawa, and Kiani 2019).

Assessing Multisensory Integration in Immersive Virtual Reality by Leveraging Haptic Stimulation

As we can measure psychophysical reactions using IVR, can we use it for multisensory integration studies? In the past, virtual reality (VR) consoles relied solely on vision and stereoscopic lenses to create a sense of depth and immersion. Nowadays, with the integration of head tracking, head-mounted displays and computer-mediated visual information, IVR technology taps into our sense of body in space (proprioception). This makes IVR a crossmodal device that can simultaneously engage two or more sensory modalities (Marucci et al. 2021). Crossmodal stimuli often induce multisensory integration, a synergistic interaction and fusion of sensory information (Stein and Stanford 2008).

To Assess multisensory integration we need to typically evaluate the effectiveness of crossmodal stimulus combinations compared to their individual components in evoked responses. This evaluation can result in either multisensory enhancement or depression, depending on whether the crossmodal event heightens or diminishes event saliency (Stein and Stanford 2008). This heightened saliency or increased neuronal firing can also lead to faster response times. For instance, a seminal study on assessing multisensory integration discovered that bimodal and trimodal stimuli, which combine auditory, visual, and tactile cues, produced quicker response times than 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). These findings primarily hinge on the integration of sensory modalities and the synchronicity of the stimuli presented. However, sensory integration relies not only on stimuli and synchronicity but also on congruency.

When we see a ball falling in our left hand, we naturally expect to feel the touch ball in the same hand and not in the right. This seamless congruency of sensory inputs from different modalities is a fundamental aspect of multisensory processing. This process helps us to effectively sort and link information originating from the same event, as 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 (Laurienti et al. 2003). This phenomenon underscores the importance of contextual consistency in facilitating efficient multisensory integration.

To quantify both the overall and congruency integration of crossmodal stimuli, I will use a well-known model for assessing it. The Race Model and the Coactivation Model have been proposed to explain the observed speedup in response timto 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 suggests that crossmodal stimuli can elicit faster reaction times than the fastest single stimuli.

As the standardization of experimental setups in IVR has become increasingly documented (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 aim to address the following questions: Can IVR and haptic gloves be utilized to investigate multisensory integration? How do the findings from such an experiment compare to existing literature on RMI?

We implemented an IVR-memory task wherein participants were instructed to memorize a 2D prompt indicating the correct position for placing a ball within a template. The duration for memorization was set at 3 seconds. Subsequently, the trial commenced with the appearance of the virtual ball and a wooden holder template placed in front of the participants on the virtual-environment. Participants then manipulated the virtual ball and placed it in the correct position within the template.

The haptic gloves provided feedback or no feedback in either a visually-haptic congruent or incongruent manner. To ensure balance across experimental conditions, the virtual ball was randomly placed an equal number of times in both the left and right hands of the participants.

I conducted an analysis of variance (ANOVA) to assess mistake rates and compare reaction times (RT) between different experimental conditions. Additionally, to gain further insight into the temporal and contextual aspects of multisensory integration, I employed a well-known RMI analysis. I compared the obtained results with existing literature. This involved utilizing cumulative distribution functions for each condition and constructing a conventional cumulative response time graph, as described by (Miller 1982).

In summary, IVR holds significant promise in cognitive neuroscience research, especially concerning multisensory integration. Its capacity to closely replicate real-life experiences, alongside its precise control over sensory stimuli and comprehensive data collection capabilities, positions IVR as a potent tool for delving into the intricacies 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 emphasis 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 ($\mu = 25.1, \sigma = 6.3$).
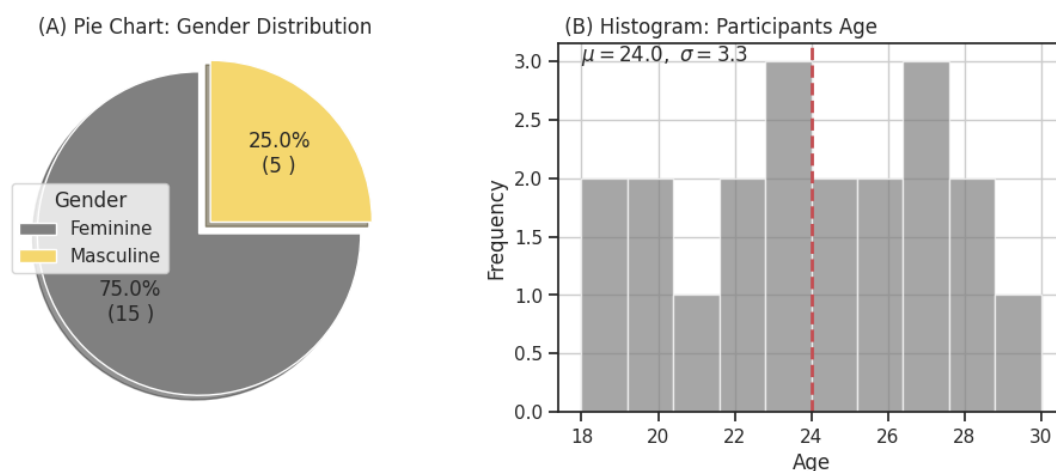

Figure 1: After cleaning the data the study included 20 participants (μ=24. years, σ = 3.3). As self-reported, all participants had normal vision, and no history of hearing loss or mental or neurological disorders, as self-reported. Notably, more women (15) responded to the call compared to men (5).

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

The participants in this study engaged in a series of tasks, each designed to elicit specific responses and behaviors. It is pertinent to mention that not all tasks from the original study were utilized in this analysis. For the sake of clarity and focus within the thesis, tasks that were performed but not included in the analysis will be referenced in the folling footnote[1]. All participants completed all questionnaires and tasks. The following description provides an overview of all tasks analyzed:

Questionnaires: 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 Eval-

---

[1]Before the IVR experience, participants completed the Edinburgh Handedness Questionnaire to determine their handedness.Additionally, participants performed a one-minute heartbeat count task before and after the IVR task.

uation Questionnaire, designed to gather their perceptions of immersion, particularly considering tactile stimuli.

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. 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.
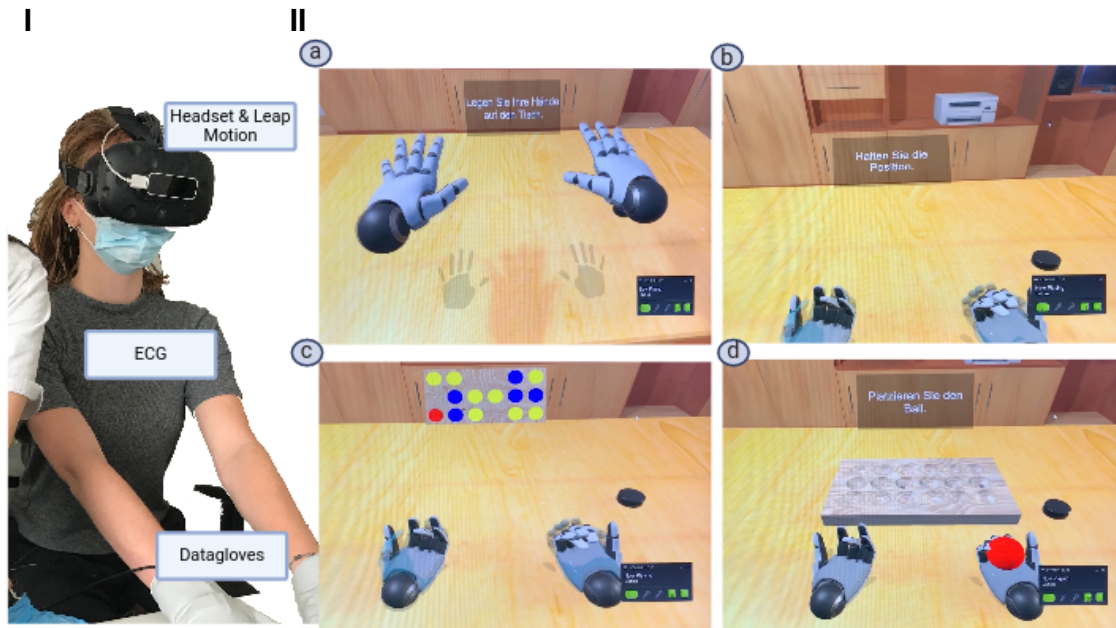
Figure 2: IVR Memory-Motor Task: Panel I shows a participant wearing equipment. Panel II shows the different stages of one trial. (a) participants first view. Here they are asked to place the hands in the virtual table to start the calibration process. 2 seconds after placed the screen changes; (b) the 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 for three seconds; (d) red ball is introduced in either hand an equal amount of times. Also the 3D template for ball placement ppears on the table 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.

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 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

Descriptive Statistics of the 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 Preprocessing and Analysis to Test Haptic Glove Condition

Response times were calculated based on the interval between the virtual ball entering the scene and its disappearance. Trials where the error button was pressed were excluded from the analysis. Prior to analysis, outliers were identified and corrected by removing data points deviating

more than 3 times the median absolute deviation (MAD), equivalent to 3 standard deviations assuming a normal distribution (Innes and Otto 2019). No response times were excluded on account of being too fast; however, 13% of trials were excluded due to excessive slowness. The final dataset comprised 1826 response times, approximately 24 per condition per participant. Following outlier removal, 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.

Cumulative Distribution Function 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:

$$F_y(t) \le F_x(t), \quad t > 0, \quad (1)$$
$$F_z(t) \le F_x(t), \quad t > 0, \quad (2)$$

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 ($V \ne T$). 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.

3.Results

Comparison of Cybersickness Levels Before and After Task, Alongside Subjective Evaluation Descriptive Statistics

We conducted a paired-sample t-test to compare cybersickness levels before (Pre) and after (Post) the Virtual Reality Experience. Significant difference was found between Pre (M=1.10, SD=0.12) and Post (M=1.18, SD=0.19) conditions; $t(21) = -2.65$, $p = 0.015$, indicating symptoms transitioned from "not present" to "somewhat present".

Nineteen respondents answered 27 questions on the Virtual Reality Subjective Evaluation Questionnaire, rating them from 1 (Does Not Apply) to 7 (Totally Applies).

The haptic gloves were perceived to enhance the experience, with question 24 scoring "Applies" for perceived immersion (Mdn $= 6$, $\mu = 6$, $\sigma = 0.76$), and question 19 also rated as "Applies" for task enjoyment (Mdn $= 6$, $\mu = 5.4$, $\sigma = 1.04$).

However, haptic feedback was perceived as having a "Neutral" effect on performance (question 26: Mdn $= 3$, $\mu = 3.6$, $\sigma = 1.49$), response time (question 12: Mdn $= 3$, $\mu = 3.7$, $\sigma = 1.74$), and results (question 7: Mdn $= 6$, $\mu = 6$, $\sigma = 0.76$).

The memory task was generally perceived as easy, with disagreement on the difficulty of remembering the red ball's position (question 8: Mdn $= 5$, $\mu = 5.1$, $\sigma = 1.37$).

Question 13 about haptic feedback possibly eased ball placement showed high variance(Mdn $= 5$, $\mu = 4.2$, $\sigma = 1.79$), then the "rather applies" is to be taken carefully.

Overall, participants reported increased immersion and enjoyment with the added gloves, but no perceived performance enhancement. Variations in responses to the last two questions suggest uncertainty about task ease. For detailed breakdown, please refer to the supplementary section.

Examining the Impact of Haptic Glove Stimuli on Response Time

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.
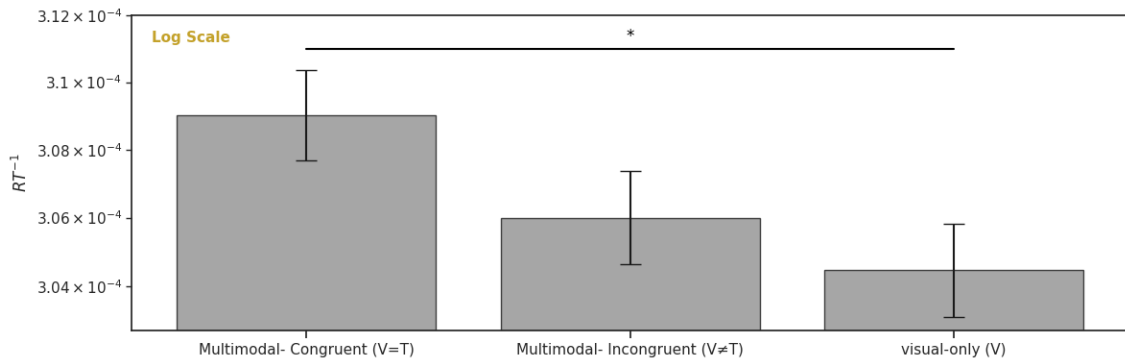

Figure 3: Passive haptic stimulus effect on RT. Boxplot of a One-way repeated-measures ANOVA using RT^{-1}. The test revealed a significant main effect for the conditions ($F(2, 38) = 3.4$, $p \leq 0.043$, $\eta p^2 = 0.15$).

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.

Analyzing Cumulative Distribution Functions and Race Model Inequality

In this section, I utilize the Cumulative Distribution Function (CDF) from the Race Model Inequality (RMI) model to evaluate the significance 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.

4.Discussion

In this study, I examined whether an experimental setup using IVR and haptic gloves could reproduce established practices in multisensory integration. I capitalized on the discrepancy between haptic gloves and real-life experiences to investigate the impact of congruency on multisensory integration. Specifically, I focused on psychophysical experimental setups that evaluate Miller's

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: Each row is organized around a percentile estimated for the cumulative distribution function (CDF). The estimations were made for each participants RT indiviually and then grouped together around percentile. The column "Time Range (ms)" shows the fastest and slowest RT for the percentile group. The column "difference" indicates the difference between the minimal and maximal response time. Both conditions, F_y (bimodal-incongruent) and F_z (bimodal- congruent) are compare to visual-only G_x in a single-side t-test.

inequality. The findings not only confirm the utility of IVR in studying multisensory integration but also offer a pathway to merging insights from psychophysics and cognitive research into a unified framework.

Furthermore, preliminary results prompt inquiry into the role of higher-order congruency multisensory integration, such as "feeling touch where we see it," in early integration processes extending beyond temporal and spatial congruency integrations.

To validate the construction of stimuli, I analyzed and modelled the reaction times across congruent, incongruent, and visual-only conditions. In total, more than a third of the trials yielded responses within the narrow timeframe of 3100 to 3200 milliseconds from the onset. While the ANOVA results unveiled a significant overall difference among all conditions, attempts to pinpoint specific pairwise distinctions using a Tukey HSD test yielded no significant findings. However, a GLMM analysis revealed a notable disparity between the unimodal visual-only condition (V) and bimodal congruent condition ($V = T$) conditions. Interestingly, no significant differences were discerned between the bimodal congruent condition ($V \neq T$) and visual-only (V) conditions.

When conducting a t-test, the RMI analysis reveals a significant discrepancy in the cumulative distribution functions (CDF) between the unimodal visual-only condition (G_x) and the bimodal congruent condition (F_y). Specifically, for F_y, this t-test identified a significant difference compared to the visual-only condition within the narrow time frame of 3100 to 3200 milliseconds (the initial 100 milliseconds as shown in Figure 4 panel b), representing the period of highest response density. Notably, no significant differences were observed for the incongruent condition across any period.

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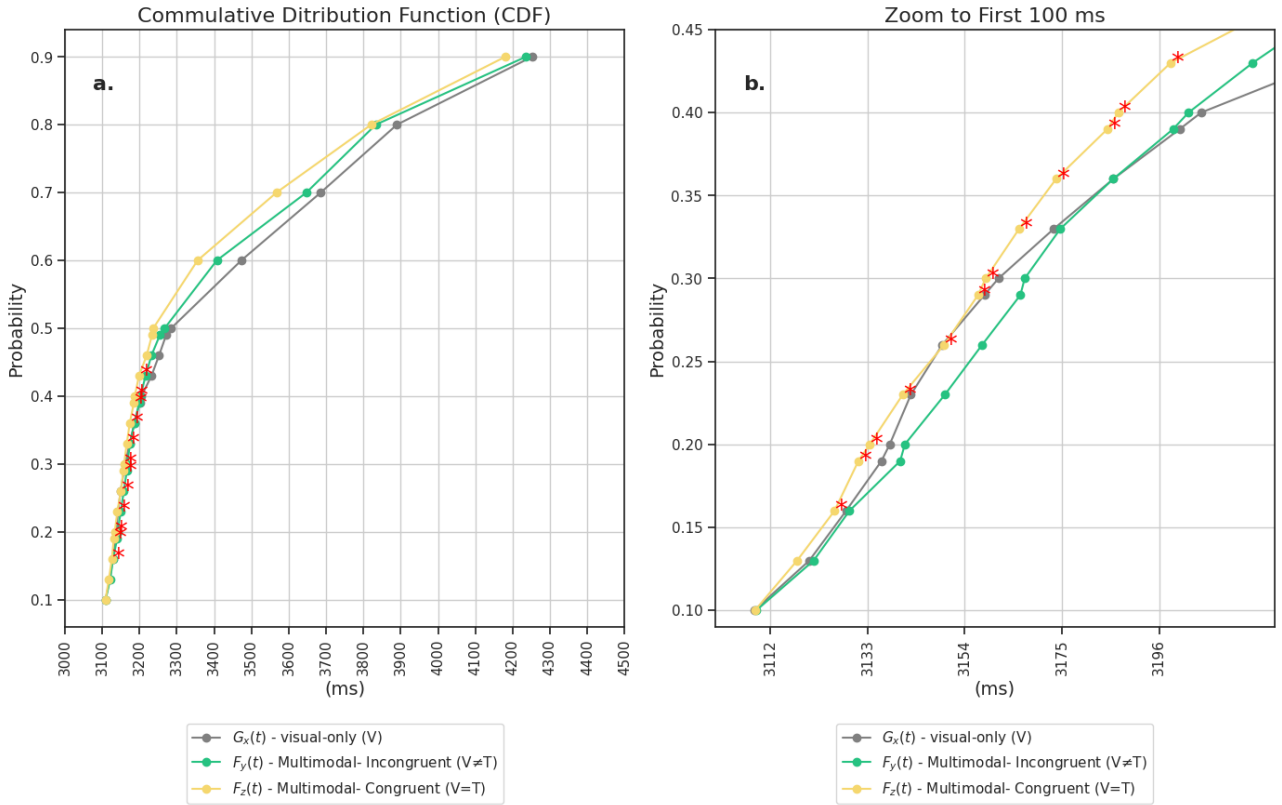
Figure 4: Cumulative distribution functions F(t) of RTs for F_y (bimodal-incongruent) and F_z (bimodal- congruent) and visual-only G_x. (a.) Displays the estimated percentile points for each of the three functions of interest: G_x, F_y, F_z, considering all participants. (b.) Is a zoom in the first 100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

condition ($V \neq T$) and the visual-only condition (V), where within the first 100 ms of the complete RT range, the unimodal 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, like in previous literature, again raises the question of a possible violation of the RMI model. Particularly as a bimodal condition ($V \neq T$) demonstrates faster responses compared to an unimodal condition (V).

In the study by Swinkels, Veling, and Schie 2021, the researchers investigated the embodiment illusion, which involves the reported feeling of seeing one's body where it is not. They found similar results to your previous study regarding the insignificance of incongruent visual-touch stimuli for reaction times (RT). Their hypothesis proposes that the lack of a difference in the Race Model Inequality (RMI) between congruent and incongruent conditions may be attributed to the simultaneous occurrence of visual and vibrotactile stimuli. This simultaneous presence could potentially facilitate multisensory integration, irrespective of whether this crossmodal stimulus is functionally linked at a higher level (Swinkels, Veling, and Schie 2021).

Furthermore, another relevant study explored the impact of multisensory integration in virtual reality (VR) on task performance, such as hitting targets and duration. This study found that only in the visual-touch condition was there an increase in performance, although this improvement was not reflected in P300 responses (Marucci et al. 2021).

These findings suggest that the congruency of visual and tactile stimuli may enhance performance in certain tasks within VR environments, but the underlying neural processes may not necessarily correspond to traditional measures like P300 responses. Additionally, my results point towards an early effect (within 100 ms) that later dissipates into integration following the redundancy effect. This might have been the case for findings in Swinkels, Veling, and Schie 2021.

I speculate that the data shows that the incongruent effect is outweighed by integration caused by the synchronicity of stimuli and sharing the same receptive field. It would be interesting to look further into this phenomenon in future research especially since we remain largely ignorant of the areas of the human brain sensitive to semantic or contextual congruence between multiple sensory cues.

Limitations

Some of this study's limitations come from the fact that each trial in our IVR task spans a longer time frame when 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) as I addressed it.

Another matter of care is the integration of all interfaces (e.g., ECG, Gloves, HMD), which should be done following the new proposed standard (Wiesing, Fink, and Weidner 2020). Especially considering that the differences between somatosensory conditions might be as small as 20 ms and latency integration issues over 7 ms might hinder the results.

Additionally, previous research on RMI employs various methods to assess the significance of differences between CMF conditions, yet more consensus on the most appropriate methods is needed.

An finally, future studies could benefit from larger sample sizes, such as involving 60 participants, to draw more definitive conclusions regarding the integratory enhancement effect of congruent stimuli, as the time effect on reaction time might be as small as of 20 ms. In the future the programming of the experiment should give the researcher different cutoff moments (e.g. movement initiation vs trial completion) to have better control over the RT skewness and variance, thus avoiding the skewness problem and increasing the significance of the results.

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.

Marucci, Matteo et al. (2021). "The impact of multisensory integration and perceptual load in virtual reality settings on performance, workload and presence". In: *Scientific Reports* 11. URL: https://api.semanticscholar.org/CorpusID:232089107.

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.

Zhang, Zhuang et al. (2023). "Active mechanical haptics with high-fidelity perceptions for immersive virtual reality". In: *Nature Machine Intelligence* 5, pp. 643–655. URL: https://api.semanticscholar.org/CorpusID:258973937.
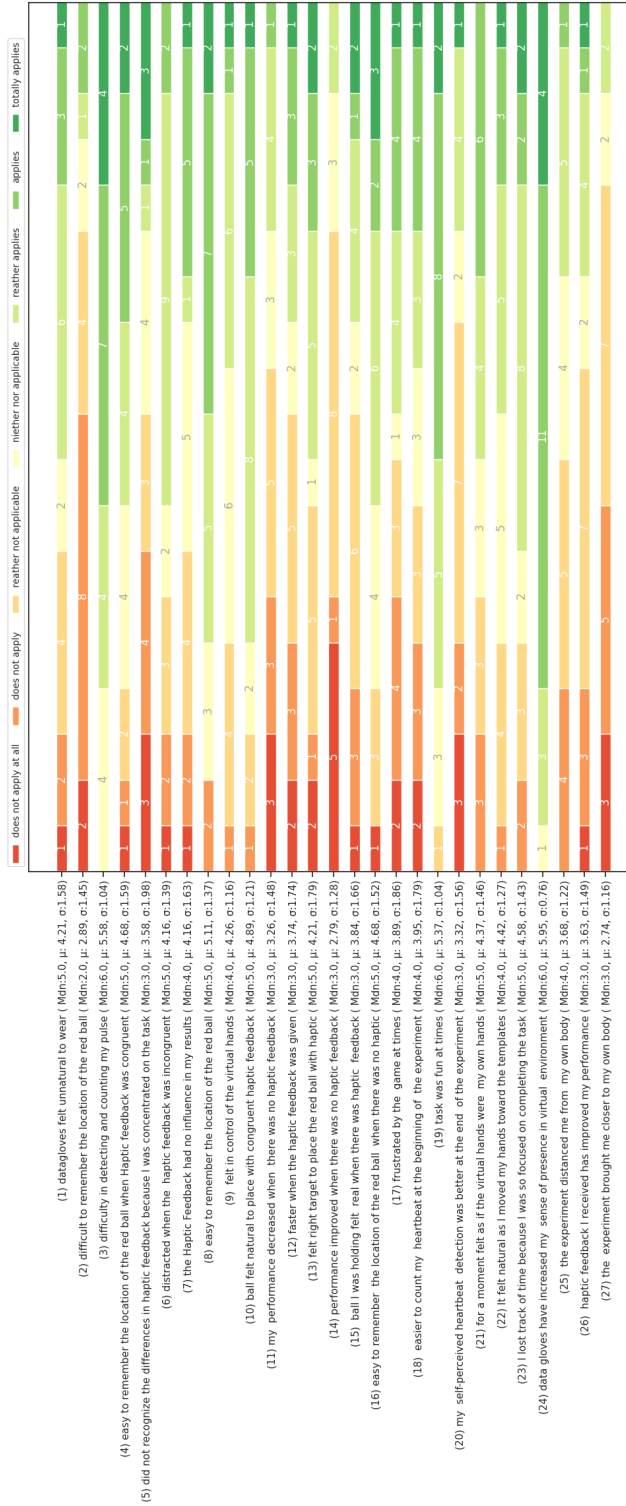
Additional Material

Figure 5: Results: Virtual Reality Subjective Evaluation Questionnaire