University of Technology Delft

Master Thesis

Msc Robotics

Exploring the Influence of Visual, Auditory, and Mental Stimuli on Cognitive Load in a Virtual Shopping Environment

Name:	J.L.A. Mulder
Student number:	4538323
Supervisors:	S. L. Cucinella
	Dr. ing. L. Marchal Crespo
	Prof. Dr. Ir. J. C. F. De Winter

February 27, 2024

Preface

I am excited to present my Master's Thesis, which investigates the impact of visual, auditory, and mental stimuli on cognitive load in a virtual shopping environment. The task of each participant was to read items from a shopping list, find the products on the shelf, and collect them in the cart in a virtual supermarket environment under seven different conditions. These conditions ranged from baseline tasks to variations with added visual, auditory, and cognitive stimuli. Various measures, including subjective NASA-TLX ratings, performance metrics, heart rate, heart rate variability, hand movement smoothness, head stillness duration, gaze behavior, and skin conductance. The findings of this research offer perspectives on the effects of varying levels of visual, auditory, and mental stimuli on cognitive load. Additionally, this study provides insights into how the stimuli affect various conventional and underexplored measures, which may be suitable for assessing cognitive load.

In addition to the challenges of conducting this research, I faced an unexpected obstacle due to a concussion. This condition made planning particularly difficult, as the effects of the concussion were highly unpredictable and frequently impacted my concentration. Interestingly, this experience also provided unique insights into the project. I learned that patients undergoing neurorehabilitation might experience overstimulation more readily than healthy individuals. Although I do not suffer from the cognitive deficits they are experiencing, the concussion often leads to overstimulation in environments such as crowded rooms or under bright lights. Consequently, this situation, while challenging, offered me an understanding of the sensations that these patients might feel, thereby enriching my understanding of the topic which motivated me in the end.

This Master's Thesis is submitted as one of the requirements for the Master Robotics at the Mechanical Engineering Faculty at the University of Technology, Delft. The presented research was supervised by Dr. Ing. L. Marchal Crespo, Prof. Dr. Ir. J. C. F. De Winter, and S. L. Cucinella, in the Department of Cognitive Robotics in the MLN Lab.

Project code can be found on two different Github Repositories:

Unity Project: `https://github.com/jobmulder98/master-thesis-unity.git`
Python Code: `https://github.com/jobmulder98/master-thesis-python.git`

All data used in this thesis are stored at the Project Storage of the University of Technology, Delft:

TU Delft Project Storage Drive: `file://tudelft.net/staf-umbrella/MasterThesisJobMulder/`

J.L.A. Mulder
Delft, February 2024

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisors, Laura Marchal-Crespo, Joost de Winter, and Salvo Cucinella. I am particularly thankful to them for their understanding and for allowing me the necessary time to recover from my concussion. I am sincerely grateful for their support, flexibility and patience during this period.

Specifically, I express my appreciation to Laura for her enthusiasm and invaluable guidance throughout the literature review and master thesis. I am consistently impressed by her ability to provide insightful feedback with minimal information, and her critical attitude towards my work. Additionally, I would like to acknowledge Joost for his critical thinking, encouragement, and support. Our weekly meetings left me challenged, full of new ideas, and with positive energy. I am grateful for his willingness to always provide guidance whenever I sought assistance. Lastly, I extend my thanks to Salvo for always willing to help me. His positive energy was indispensable, and each meeting with him transformed my questions into ideas for the upcoming week. His creative thinking, combined with his extensive knowledge, makes him as an exceptional daily supervisor.

I also want to express my gratitude to all people from the MLN Lab who took time to support me at various stages of the thesis. Their willingness to help has been invaluable. Lastly, I would like to thank all participants who participated in the experiment.

Contents

1

Scientific Paper

Exploring the Influence of Visual, Auditory, and Mental Stimuli on Cognitive Load in a Virtual Shopping Environment

J.L.A. Mulder, S.L. Cucinella, J.C.F. de Winter, L. Marchal Crespo

***Abstract*—Immersive virtual reality (IVR) with head-mounted displays (HMDs) is expanding in various fields like training, but its effects on cognitive load from visual, auditory, and mental stimuli in virtual environments remain uncertain. This is particularly relevant in neurorehabilitation, where patients may suffer from training in overstimulating environments due to cognitive impairments. This study further explores how low and high levels of visual, auditory, and cognitive demands affect the cognitive load. Twenty-two participants used an HMD for a virtual shopping task, which involved selecting listed products and placing them in a cart, under baseline (the task without additional demands) and two stimulus complexity levels (low and high) for visual, auditory, and mental demands. The study assessed cognitive load using conventional (heart rate, variability, skin conductance, performance, self-reported questionnaire) and underexplored measures (head stillness, hand smoothness, gaze behavior), and explored behavior changes due to stimulus impact. Results showed that visual and auditory stimuli had minimal effects on cognitive load, with only specific measures showing any notable differences from the baseline. Mental stimuli significantly impacted cognitive load, with high mental tasks notably affecting the measures and behavior, whereas low mental tasks showed fewer changes. This research concluded that mental stimuli significantly increased the cognitive load in virtual environments, more than visual or auditory stimuli, suggesting future virtual reality designs should prioritize managing mental load. Furthermore, the study highlights the effectiveness of head stillness and gaze behavior as innovative measures for evaluating cognitive load.**

***Index Terms*—Cognitive Load, virtual reality, psychophysiology, human factors, behavior, realism**

I. INTRODUCTION

In recent years, the use of immersive virtual reality (IVR) with Head-Mounted Displays (HMDs) has increased in many fields, including neurorehabilitation [1]. Neurorehabilitation facilitates recovery from neurological injuries [2], where patients perform task-oriented [3], highly repetitive [4], and intensive training [5] to regain their lost abilities. In this context, HMDs offer an interactive and stimulating environment [6], and increase patient motivation during training [7], which, as a result, can improve task repetitiveness [7]. However, the optimal adjustment of virtual environments, particularly how the levels of visual, auditory, and cognitive demands can affect training, remains unclear.

On the one hand, there is the assumption in neurorehabilitation that creating a realistic environment contributes to enhancing the user experience and facilitating better transfer of learning to real-world practice [8], [9]. Realism refers to the degree to which the virtual environment resembles the real world [10], and includes visual, auditory, and tactile feedback, which help to recreate real-life experiences [9]. Moreover, the higher the level of realism, the more it has a positive impact on presence [8], which tells us how much people feel present in the virtual world [11]. This can affect how well treatments work, and how they influence a person's real-life actions [12].

On the other hand, a realistic environment may overstimulate users, especially those with cognitive deficits, who may experience cognitive overload more readily than healthy individuals [13], [14]. The cognitive load refers to the level of mental resources used by a human to perform a task [15]. According to cognitive load theory, people should not be excessively stimulated; rather, they should face an adequate challenge to maintain spare capacity [16], [17]. Therefore, when designing training environments that incorporate various types of stimuli, it is important to balance the level of realism to avoid cognitive overload and ensure the effectiveness of the training outcomes.

Previous research explored how visual (moving characters), auditory (background noise), and mental (arithmetic task) demands influenced the cognitive load in an IVR shopping task [18]. The virtual supermarket was chosen since it entailed these three types of demands; supermarkets can be noisy and contain visual stimuli like people walking around. Additionally, shopping entails cognitive challenges, such as remembering various products and their prices. The study demonstrated that the addition of a secondary arithmetic task contributed more to an increase in cognitive load than visual and auditory stimuli.

However, the study only examined the impact of one level per type of stimuli on cognitive load, while certain stimuli may elevate cognitive load more than others. For instance, characters casually moving in the background may potentially increase cognitive load less than characters actively seeking interaction and making eye contact. This is because eye contact is known to directly distract humans and can even heighten cognitive load [19]. Similarly, in the case of auditory stimuli, background noise might be more easily filtered out compared to a loud conversation [20]. Moreover, in the study of [18] participants were told to finish a shopping list as fast as possible, while the response to visual, auditory, and mental stimuli may be different when they are allowed to determine their own pace. We see this, for example, in driving where visual demand leads to reduced speed to avoid becoming overloaded [21]. Also, time pressure can create tunnel vision, potentially focusing more narrowly on the central target and disregarding distractors [22]. Therefore, this study investigates the impact of different levels of visual, auditory, and mental stimuli (a cognitive, secondary task) on cognitive load in a self-paced immersive virtual shopping environment.

In this study, various measures associated with cognitive load were used. These included task performance, where performance tends to decline with increasing cognitive load [23]. Additionally, subjective assessments were conducted using the NASA-TLX questionnaire [24], where higher scores correspond to higher perceived cognitive load. Psychophysiological measures such as skin conductance, heart rate, and heart rate variability were also used. Skin conductance tends to increase with higher cognitive load [25]–[28], while heart rate increases and heart rate variability decreases with higher cognitive load levels [29], [30].

Besides the conventional measures, this research incorporates underexplored measures. Cognitive demands have been found to affect the speed and variability of hand movements in [31]. When participants are deeply engaged in selecting and placing products into a cart, this heightened concentration is hypothesized to have an impact on the speed and trajectory of hand movements. Therefore, hand movement smoothness, which can address changes in hand movements, was used as a measure. Furthermore, when cognitive load increases, participants may focus on one point to concentrate and minimize their head movements. This phenomenon also occurs with the eyes, where the fixation rate decreases for a higher cognitive load [32], [33]. Therefore, the total duration during which the head remains still was used as a measure. Lastly, an increase in cognitive load can lead to a shift in oculomotor behavior [34], potentially leading to longer total time spent looking at specific areas of interest. For instance, adding visual stimuli to the shopping task may potentially increase the cognitive load and distract participants, resulting in a gaze pattern change. Hence, the total gaze duration on areas of interest within the virtual environment served as a measure.

The results of this study can provide insights into how different levels of visual, auditory, and mental stimuli impact the cognitive load. Moreover, this study provides insights into how the stimuli affect various conventional and underexplored measures, which may be suitable for assessing cognitive load.

II. Methods

A. Participants

The study was approved by the Human Research Ethics Committee of the Delft University of Technology. In total, twenty-two (11 females, 11 males) individuals who were recruited via word-of-mouth performed an experiment. All participants were students (all university and higher professional education) in the age range of 23 to 27 years. Participants were excluded if they had ever suffered from an epileptic attack, if they had used benzodiazepines in the past month, if they had any motor/cognitive disorders, or if they had experienced anxiety or motion sickness before when using an HMD. Prior experience with IVR was not a requirement. All participants provided explicit written consent to participate in the experiment and were rewarded with a €15,- gift card.

B. Experimental Equipment

The experiment was conducted in a room with enough space for the participants to freely move, without causing collisions. The experiment included an HMD (HTC Vive Pro Eye, HTC Vive, Taiwan & Valve, USA) with a field of view of 110 degrees and Dual OLED screens with a resolution of 1440 x 1600 pixels per eye (2880 x 1600 pixels combined), and one HTC Vive tracked controller (HTC VIVE Pro Controller 2.0). The HTC Vive Pro Eye featured built-in eye-tracking technology and recorded data at a frequency of approximately 90 Hz, dependent on the Unity processes running.

The virtual environment was developed with the Unity game engine (Unity Technologies, USA), version 2021.3.11f1. The SteamVR software (version 2.7.3., Valve Corporation, USA) was used to connect the HMD with Unity. For eye tracking, the TobiiXR SDK and SRanipal SDK packages were downloaded. The environment was obtained from [18], consisting of the supermarket building with products, shopping list, shopping cart, and shelves with products, from the Unity Asset Store. Additionally, the Low Poly Animated People package was used (Polyperfect, USA) for the avatars. Furthermore, audio was acquired from the Envato Elements store (elements.envato.com) and Pixabay (pixabay.com). The HTC Vive Pro Eye headphones were used to play the audio fragments. The screen of the virtual environment was captured with the Windows screen recorder (XBOX Recorder, 60 Hz).

The Biotrace (MindMedia, NeXuS-4) was used with a Bluetooth connection to obtain the electrocardiogram (ECG) and electrodermal activity (EDA) signal (ECG; 1024 Hz, 24-bit. EDA; 32 Hz, 24-bit). A Lead II configuration was used for electrode placement of the ECG. The electrodes for the EDA signal were attached to the index and ring finger of the left hand. Data was obtained via MindMedia software (Biotrace+ NX4, 64 bit).

C. Experimental Task

The experimental task required participants to read products from a shopping list and memorize them, pick up the products from the shelves, and place them in the cart. The list had to be completed in chronological order, consisted of 30 unique items, and was consistently displayed in the upper right corner, positioned above the shopping cart, as shown in Figure 1a. To collect the products, participants were able to freely move around in the environment. Participants had to hold the button at the back of the controller to grab an object. When the participants released the button, the product was released. The participants were not able to see their bodies, but an avatar representing the right hand of the participant was visible in the virtual environment, as illustrated in Figure 2.

Participants were required to perform seven conditions. After 122 seconds, the condition stopped automatically. The name and physical appearance of the products remained the same in each condition. However, the order of products on the shopping list and the location of the products on the shelves differed between conditions. Yet, the first twenty items on the list followed a non-random logical pattern, which was designed to appear random to the participants. This adjustment

(a) The task environment from a remote perspective, including the shopping list, the cart, and the main shelf with products.

(b) The task environment from a remote perspective during the visual high condition with characters surrounding the main shelf.

Fig. 1: The task environment from a remote perspective.

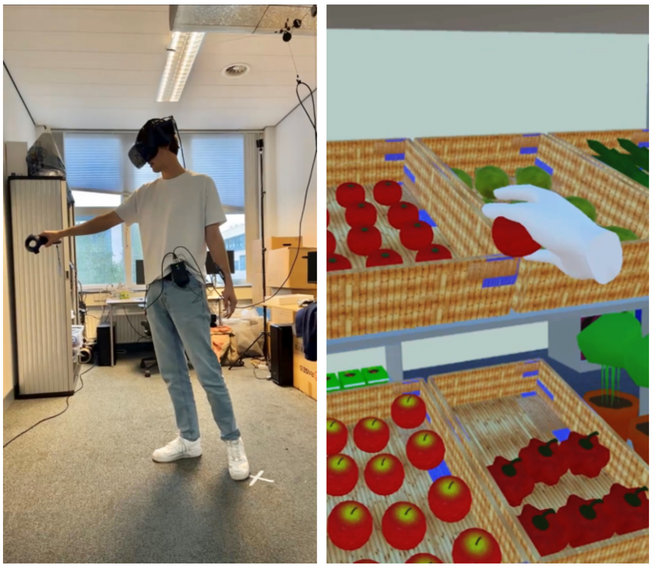


Fig. 2: Left; A participant wearing the HMD, holding the controller in the right hand, reaching for a product. Right; The environment that the participant perceives while grabbing a product.

ensured approximately equal distances from the products to the shopping cart across all conditions. By doing so, potential biases in measuring the items collected per minute were minimized.

D. Experimental Conditions

The experiment consisted of seven different conditions:
1 - Baseline: The baseline condition consisted only of the experimental task. Hence, there were no extra stimuli present, i.e., no audio, no extra visuals, no secondary task.
2 - Visual Low: There were eight idle characters (standing in the environment), and 22 moving characters (spawning, moving along a trajectory, and disappearing). Eight out of these 22 moving characters were action-moving (performed an interaction action towards the participant, such as waving or a thumbs-up gesture). The visual low condition represented the supermarket from daily life. Characters were casually dressed, actions were performed behind the main shelf, and the characters in the same aisle were all idle and were located at least four meters away from the starting point of the participant. For an overview of the condition characteristics, see Table I, more detailed information can be found in Appendix B.2.1.
3 - Visual High: There were eleven idle characters, and 44 moving characters, 28 of which were action-moving. Distinctions between the low and high visual conditions focus on character types, actions, events, and their locations. Two alarm lights were added to the ceiling above the main shelf. Furthermore, the characters were closer, i.e. at least two meters from the participant, and had odd appearances such as police, clown, Viking, chef, etc. (Figure 1b).
4 - Auditory Low: The audio added in the auditory conditions was spatial, which means that the audio in the headphones turned along with the movement of the HMD, and the volume was dependent on the distance to the source. In this condition, there were two footstep trajectories (footstep audio along a trajectory), and three idle audio sounds (continuous audio, not moving). Footsteps were translated at a walking pace, the volume was set to approximately 85 dB, and the sound origins were at least five meters away from the starting point of the participant. For an overview, see Table II, and more detailed information can be found in Appendix B.2.2.
5 - Auditory High: Audio consisted of three footstep trajectories, five idle sounds, two idle sounds that were not continuous but repeated with breaks in between, and six incidental sounds (short fragments, two meters away). The volume was approximately 95 dB, the location of the audio was at least two meters away from the participant, and footsteps were translating at both a walking and running pace.
6 - Mental Low: A secondary task, specifically the 0-back task (a sub-task of the n-back task which is known to increase cognitive load [35]), was added to the main task. The participant heard an auditory voice that gave 4 sets of random single-digit numbers. During the 0-back task, participants were required to repeat out loud the number that they had heard 0 numbers ago, i.e., they had to repeat the number that they had just heard.
7 - Mental High: A secondary task, specifically the 2-back task [35], was added to the main task, which had the same format

TABLE I: Overview of the different factors that were used to influence cognitive load in the visual conditions, including examples.

Factor	Visual Low	Visual High
Character type	Casual characters	Odd, remarkable characters
Character actions	Mainly walking and looking at the participant	Also doing more unique moves, such as waving, dancing, etc.
Events	Only characters	Also non-character objects as visual distractors, such as a flashing alarm.
Location	Behind the main shelf, at least four meters from the participant.	Also next to the main shelf, at least two meters from the participant.
Examples	Casually dressed character walking, waving, or looking at you.	Oddly dressed characters such as clowns, firemen, farmers, Santa Claus, or pilots. Alarm flashing, characters dancing, jumping, calling.

TABLE II: Overview of the different factors that were used to influence cognitive load in the auditory conditions, including examples.

Category	Auditory Low	Auditory High
Dynamics	Static or translating at a walking pace.	Static or translating at a walking, running, or driving pace.
Volume	Approximately 85 dB.	Approximately 95 dB.
Duration	Continuous audio fragments.	Continuous and short, repeating audio fragments.
Location	At least five meters from the starting point of the participant	At least two meters from the starting point of the participant.
Examples	Checkout desk bleeps, couple arguing, crowd conversations, walking footsteps in front of the participant	In addition to the auditory low condition; Police cars with sirens driving by, a person calling to the participant nearby, running footsteps behind the participant, truck in reverse, bottles falling.

as the 0-back task. During the 2-back task, participants were required to repeat out loud the number that they had heard 2 numbers ago.

Each participant experimented in a random condition order. Since counterbalancing was not possible with 22 participants, a computer randomized the conditions using a randomization tool (http://random.org), which does not rely on random seeds.

E. Dependent Measures

This study incorporates metrics known to be linked to cognitive load, such as performance [23], subjective assessments [24], and psychophysiological measures [36]. Finally, under-explored measures were included to assess their suitability for evaluating cognitive load. The following dependent measures were used in each condition:

- *Subjective NASA-TLX:* The weighted NASA Task Load Index (NASA-TLX) was used to assess the subjective cognitive load post-task [24]. Participants rated how much the aspects *Mental demand*, *Physical demand*, *Temporal demand*, *Performance*, *Effort*, and *Frustration* affected their cognitive load on a twenty-point scale (0 to 100, in increments of 5). The weighted version of the NASA-TLX integrates subjective weights for each item. These weights reflect the perceived significance of each aspect assessed. A higher weight assigned to an item indicates that it contributes more substantially to the final score.
- *Performance:* The items collected per minute for each condition were computed by dividing the total number of items collected by the number of minutes taken to collect the final item.
- *Heart Rate and Heart Rate Variability:* The heart rate (HR) and heart rate variability (HRV) were computed by detecting RR-peaks in the raw ECG signal using the *biosppy.signals.ecg* package [37]. The heart rate was defined by the number of RR intervals per minute and the heart rate variability was defined by the root mean square of two consecutive RR intervals (RMSSD).
- *Skin Conductance Level:* Skin conductance level (SCL) was computed by analyzing the EDA signal with the *neurokit2* package.
- *Hand Movement Smoothness:* The minimum jerk model [38] suggests that the smoothness of the hand movements is executed with a trajectory that minimizes the rate of change of acceleration, also known as the jerk. Therefore, the smoothness of hand movements during object grabbing was computed by calculating the mean jerk. The measurements were performed for trajectories of the hand while grabbing an item since these movements are always point-to-point and they could be tracked with high precision.
- *Head Stillness Duration:* The total time where the filtered signal of the angular head acceleration registered values lower than an absolute value of $3\,m/s^2$ was computed.
- *Gaze Duration on an Area of Interest:* The total time spent looking at each area of interest (AOI) was computed. The AOI is determined by tagging objects in the environment as the *main shelf* (including products and shelves), *cart*, *shopping list*, or *other objects* (all other objects not assigned to the main shelf, cart, or shopping list). Participants' gaze locations were then assigned to one of these tags.

F. Experimental Protocol

The experiment took place in a destined room of the Cognitive Robotics Department of the TU Delft. Upon arrival to the experimental room, participants were asked to read and sign an informed consent. After signing the informed consent, the participants were given instructions on how to perform the experiment.

Participants were instructed to read items from the list, retrieve them from shelves, and place them in the shopping cart. They were guided on using the right controller to grab and release products with the button on its back, emphasizing that the left hand should not be used since electrodes were attached to that hand to obtain the EDA signal. Participants were directed to avoid throwing items and place them in the cart to ensure proper registration. They were informed that products would be removed from the cart after two seconds to prevent a pile of items, which could hinder the registration of products in the cart. Emphasis was placed on the task not being a race and that the list was probably too long to complete on purpose, with the primary goal of placing items chronologically in the cart. Participants were instructed that all listed items were located on the main shelf to discourage them from moving to other areas. Participants were familiarized with the environment and the physical appearance of the products. Additionally, safety precautions were explained, including warnings about potential motion sickness or anxiety, with the option for participants to withdraw from the experiment at any time.

After giving instructions to the participant, the electrodes to attach the EDA and ECG sensors were given to the participant to attach them to their body. The researcher always asked consent to stay in the room or to go outside until the attaching of the electrodes was finished. After attaching the electrodes, the Windows screen recorder was started to record the environment and audio, which was stopped after finishing the final condition. The participants were instructed on how to use the HMD, emphasizing the need for room and eye calibration between conditions. To ensure all participants were able to see all products, the height of each participant in the environment was set to 185 cm. This adjustment addressed potential visibility challenges for participants shorter than 170 cm, especially concerning products on the top shelf. Once all settings were configured, participants engaged in a practice phase, allowing them to familiarize themselves with the environment without additional stimuli or time constraints.

Following the practice trial, participants underwent the seven different conditions, with each participant experiencing a random order. After completing each condition, participants removed their HMD and filled out the NASA-TLX questionnaire online through Qualtrics (Provo, Utah, USA). Prior to each condition, both the room and eye tracker were recalibrated. Participants were not briefed about the specific condition beforehand, except in the case of the n-back tasks. For the n-back task, participants were informed that an additional task would be introduced alongside the main task. They were then given instructions on how to fulfill the task and were instructed that the main and n-back tasks had the same priority. Before putting on the HMD, an example audio of the n-back task was provided to familiarize participants with the format, and they were encouraged to ask questions if any aspects were unclear.

Upon completion of all seven conditions, participants were required to fill out an additional NASA-TLX questionnaire to determine the weights. The experiment lasted approximately 75-90 minutes per participant. For a schematic overview of the timeline of the experiment, see Figure 3.

G. Data Collection and Filtering

Data was collected using different devices, where in the end three types of data were collected for each participant. The HMD and Unity stored the collected data in a separate *.csv* file for each condition, for each participant. Screen recordings of the environment were saved in *.mp4* files using the Windows screen recorder. The data from the ECG and EDA was saved in comma delimited *.txt* files, one file per participant.

Since the file containing ECG and EDA data was recorded continuously, it required synchronization with the *.csv* files. This synchronization was achieved by aligning timestamps: one captured at the onset of ECG and EDA signal recording, and two others indicating the start and end of data collection in the *.csv* files. Using the constant frame rate from the ECG and EDA device, the range of data points within the *.txt* file, spanning from the start to the end timestamps, could be precisely delineated.

The signals for the head and head movements did not have similar frame rates and were therefore interpolated so that each time frame equaled 0.01 seconds (100 Hz). After the interpolation, the signal was filtered with a butter filter (cutoff 20 Hz, order 3). Finally, the second time derivative was calculated for the head acceleration, and the third time derivative was used to compute the jerk.

Eye recordings had to be filtered; when eyes were closed the eye tracking was not able to write reliable data. This data was interpolated at the moments when participants were blinking and the AOI tag *invalid* was used. The ECG signal was filtered and analyzed using the *biosppy.signals.ecg* package. However, the data needed further filtering since there were numerous peaks undetected, and some detected peaks were false. To filter this out, the heart rate was computed and outliers were removed using Inter Quartile Range (IQR) outlier detection [39] [40]. The IQR, calculated as the range between the 25th and 75th percentile sample quantiles (Q_1 and Q_3 respectively), equals $Q_3 - Q_1$, with data points outside the lower whisker limit ($Q_1 - 1.5IQR$) and higher whisker limit ($Q_3 + 1.5IQR$) considered as outliers. After filtering each signal, the features and graphs were extracted using Python (version 3.11). All code for filtering and post-processing can be found in the Github repository.

H. Data Analysis

In addition to the analysis of the response of the dependent measures described in II-E on different stimuli, an exploratory behavior analysis was performed. A combination of gaze behavior, performance, and interaction between the controller and the products was used to create a chart to analyze the behavior and tactics of the participants. Based on the results of the exploratory behavior analysis, three features were extracted to find differences in tactics or behavior between the baseline and the other conditions: mean *shopping list* visits, mean *shopping list* time, and gazing at *shopping list* while grabbing. These are further elaborated in subsection III-B.

The objective of this research was to investigate how different levels of types of stimuli affect the cognitive load. To achieve this, the research design incorporated participants as random effects to account for individual differences, and conditions as fixed effects to evaluate the distinct stimulus levels. Additionally, the study controlled for sequence effects

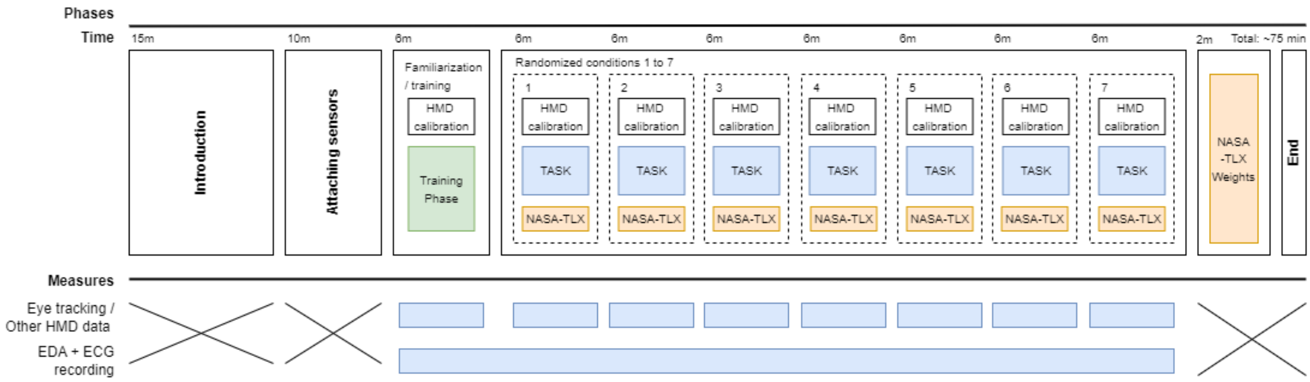


Fig. 3: Schematic overview of the experimental protocol, spanning approximately 75 minutes (with a maximum duration of 90 minutes). Data collection started with the ECG and EDA recordings at the onset of the familiarization/training phase and continued until the end of the experiment. HMD-based data collection occurred separately for each trial, with calibration of the HMD and eye tracking preceding each condition. Post-condition, the HMD was removed, and participants completed the NASA-TLX questionnaire. After the final condition, the weights for the NASA-TLX were assessed.

to account for learning over time. A Linear Mixed Effects (LME) model [41] model was used to analyze the data. The LME model is defined by:

$$\mathbf{y} = \mathbf{X_f}\boldsymbol{\beta} + \mathbf{Z_r}\mathbf{U} + \boldsymbol{\Psi}$$

Where $\mathbf{y}$ is the dependent measure, the term $\mathbf{X_f}\boldsymbol{\beta}$ is the fixed effect, the term $\mathbf{Z_r}\mathbf{U}$ is the random effect, and $\boldsymbol{\Psi}$ is the residual distribution. The model was employed for performing both the analysis of the dependent measures and the exploratory behavior analysis. The model made six comparisons; it compared all visual, auditory, and mental conditions to the baseline condition. The obtained p-values were adjusted using Holm's method [42] which is a relatively conservative correction to regulate the family-wise error rate. R-Studio (R version 4.3.1) was used to compute the p-values, which were considered significant when $p < 0.05$ after correction.

Cohen's d was used as an effect size metric, which describes the difference in means relative to the pooled standard deviation [43]. However, it did not consider the order of conditions, unlike the LME model. Hence, before computing Cohen's d, the data underwent sorting based on the condition order per participant, followed by z-score transformation, and then reverted to its original order. Cohen's d_z was also calculated and added to the appendix, which quantifies the difference between means divided by the standard deviation of the difference scores [44]. Cohen classified effect sizes as small ($d = 0.2$), medium ($d = 0.5$), and large ($d \geq 0.8$) [43]. Python (version 3.11) was used to calculate Cohen's d and d_z.

III. Results

A. Conditions Compared To Baseline

Upon analysis, the obtained EDA signal seemed corrupted for many participants since the hand movements had caused disconnection with the sensors, and therefore the filtered data was unreliable. For this reason, skin conductance was excluded from the study. Table III shows the results of the statistical analysis for the remaining measures. The values represent the effect size, and significant differences are highlighted in bold. Due to space constraints, specific metrics such as means, standard deviations, skewness, kurtosis, and Cohen's d_z are not directly provided in the table but are available in Appendix A.2 and A.9. The largest correlations (Figure 4) were found between the gaze duration on the *main shelf* and the gaze duration on the *shopping list* ($r = -0.72$), between the mean heart rate and mean heart rate variability ($r = -0.68$), and between the performance and the head stillness duration ($r = -0.62$). Box plots of the gaze duration on *other objects*, performance, and weighted NASA-TLX scores are shown in Figure 5, 6, and 7 respectively.

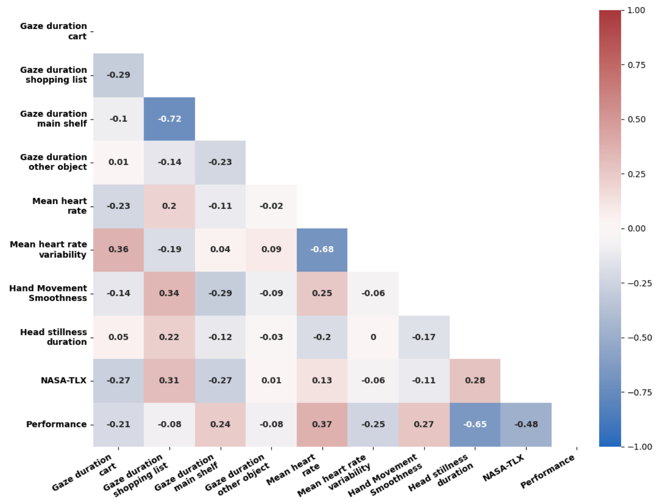


Fig. 4: Correlation matrix of the dependent measures.

Regarding the visual low and high conditions, the gaze duration on *other objects* in both the visual low ($d = 1.32$, $p = 0.049$) and visual high ($d = 2.08$, $p < 0.001$) condition was significantly longer relative to the baseline. The auditory low condition shows no significant differences w.r.t. baseline, and the auditory high condition shows a significantly larger difference w.r.t. baseline for the NASA-TLX ($d = 1.07$, $p < 0.001$). The mental low condition shows significant

TABLE III: Results from linear mixed effects model. The reported effect size Cohen's d is highlighted in bold text if the corrected p-value was significant ($p < 0.05$). The italic names in the gaze duration measures represent the different areas of interest in the environment, as described in subsection II-E.

Measure	**Visual Low**	**Visual High**	**Auditory Low**	**Auditory High**	**Mental Low**	**Mental High**	**Unit**
Gaze Duration on *Cart*	0.01	-0.33	-0.31	-0.18	-0.70	-0.52	$seconds$
Gaze Duration on *Shopping List*	0.05	-0.08	0.24	0.15	**0.74**	**1.31**	$seconds$
Gaze Duration on *Main Shelf*	-0.48	-0.72	-0.18	0.01	-0.59	**-1.48**	$seconds$
Gaze Duration on *Other Objects*	**1.32**	**2.08**	0.14	0.15	-0.12	0.53	$seconds$
Mean Heart Rate	0.01	-0.06	-0.05	0.05	0.27	0.20	bpm
Mean Heart Rate Variability	-0.07	-0.01	0.00	0.00	-0.30	-0.06	bpm
Hand Movement Smoothness	0.07	0.06	-0.04	0.10	0.11	0.13	m/s^3
Head Stillness Duration	0.13	-0.31	0.32	0.01	0.20	**1.11**	$seconds$
Performance	0.09	-0.26	-0.39	-0.38	-0.25	**-3.21**	$items/minute$
Weighted NASA-TLX Score	0.03	0.55	0.41	**1.07**	**1.45**	**4.82**	–

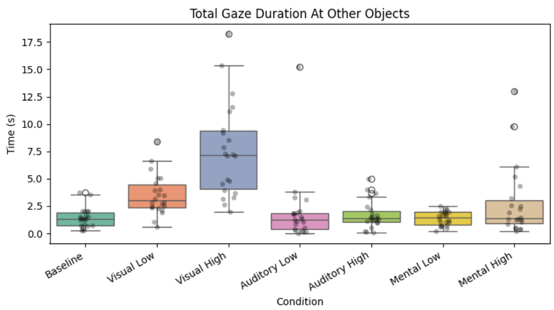


Fig. 5: Box plot showing the total gaze duration on *Other Objects* across conditions.

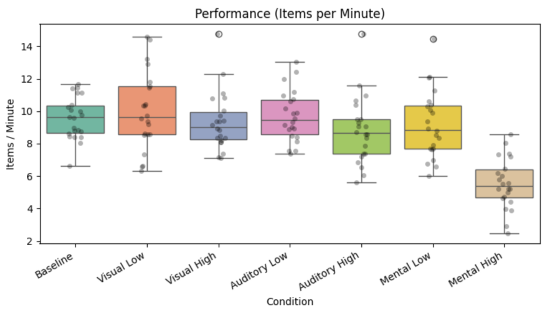


Fig. 6: Box plot with the performance, i.e. the number of items collected per minute, across conditions.

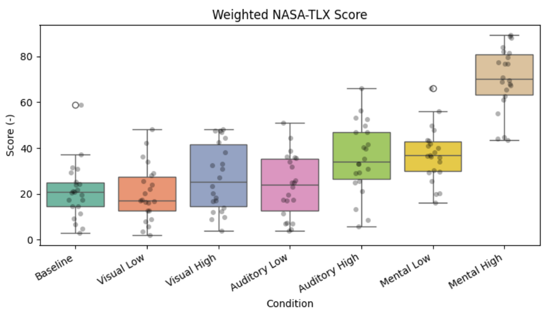


Fig. 7: The weighted NASA-TLX score on a scale from 0 to 100 across conditions.

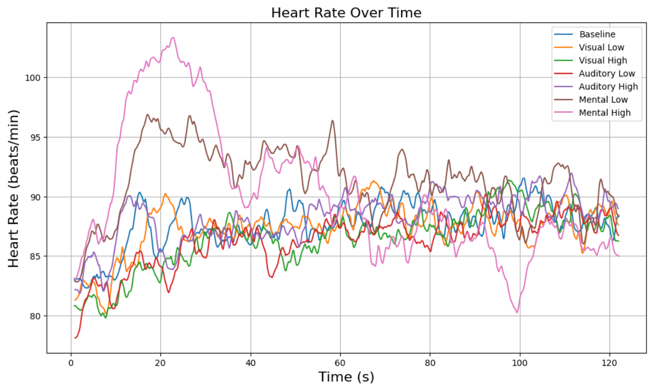


Fig. 8: The heart rate over time across all conditions, averaged over all participants.

differences w.r.t. baseline for the gaze duration on the *shopping list* ($d = 0.74$, $p = 0.038$), and the NASA-TLX ($d = 1.45$, $p < 0.001$). Finally, the mental high condition shows significant difference w.r.t. baseline for the gaze duration on the *shopping list* ($d = 1.31$, $p < 0.001$) and on the *main shelf* ($d = -1.48$, $p < 0.001$), the head stillness duration ($d = -1.57$, $p < 0.001$), the performance ($d = -3.21$, $p < 0.001$), and the NASA-TLX ($d = 4.82$, $p < 0.001$).

Upon closer examination of the mean heart rate over time, an apparent increase is observed in the mental low and high conditions during the initial 35 seconds (Figure 8). Subsequently, after 35 seconds, the mean heart rate in these two conditions returns to the baseline level. Even though this increase could be visually determined, the mean heart rate for the entire mental low and high conditions show no significant differences w.r.t. baseline condition. When using the same LME model with p-value correction for the first 35 seconds of the condition instead of the entire condition, including the participants as random effects and taking into account the order of conditions, significant differences w.r.t. to baseline were found for the mental low condition ($d = 0.70$, $p < 0.001$) and mental high condition ($d = 1.06$, $p < 0.001$).

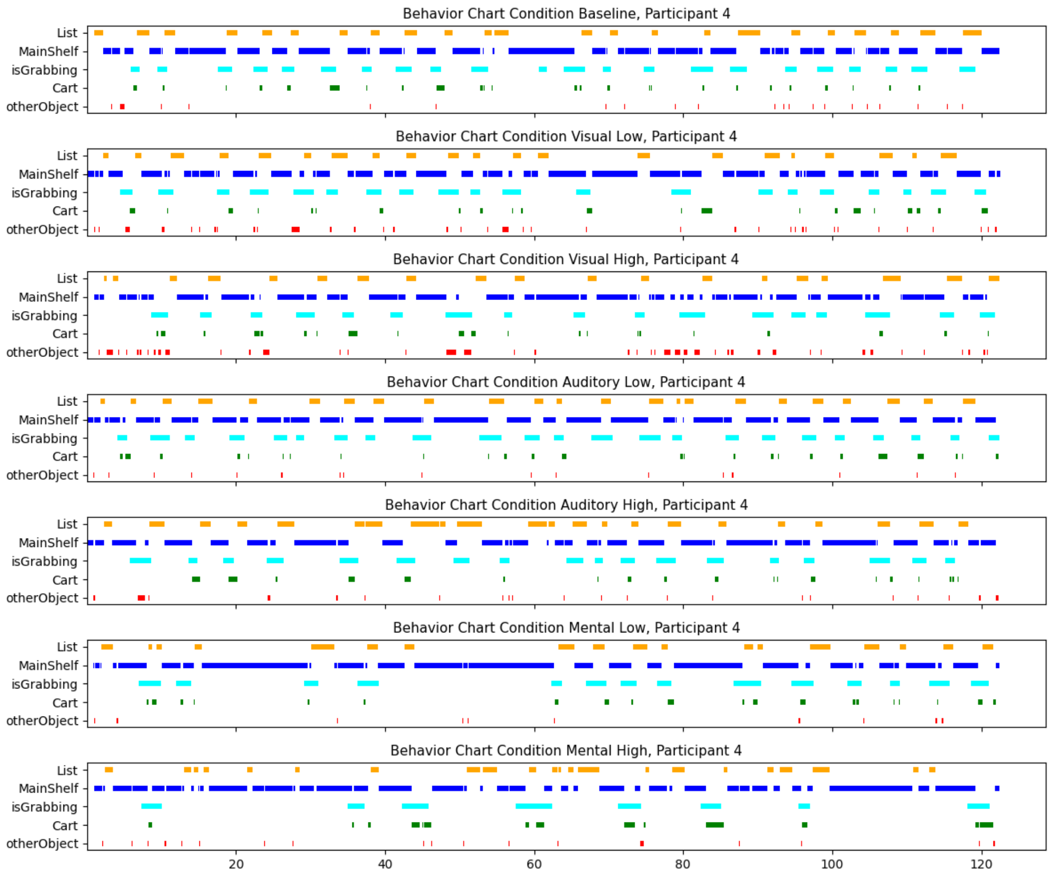


Fig. 9: The behavioral chart provides a visual representation of the gaze patterns and object interaction of Participant Four across the seven distinct conditions outlined in section II-C. On the vertical axis, the chart contains specific areas of interest within the environment, with colored bars indicating the participant's gaze allocation on a specific AOI during each condition. Furthermore, the chart includes information on instances of object grabbing. The horizontal axis corresponds to time, spanning from 0 to 122 seconds.

B. *Exploratory Behavior Analysis*

The combination of gaze behavior and interaction between the controller and the products was used to create behavior charts, to find underlying behavior patterns and tactics for each condition. An example of a behavior chart for a single participant can be found in Figure 9. The chart displays the gaze on specific areas of interest during task execution and indicates whether they were holding an item. Behavior charts for all other participants can be found in Appendix A.8.

Observations from the baseline condition revealed a strategy among participants. Initially, they consulted the list to identify the target product. Subsequently, they navigated to the main shelf to locate the product, with the duration of this step shortening if the product was found quickly. Following this, participants grabbed the object and moved towards the cart while securing the item. In this paper, this pattern is described as one routine (Figure 10). Since not all participants gazed at the cart, it was not included in the routine.

Comparing the conditions visually, some routine differences emerged. Overall, the visual low and high conditions indicated more gaze time on *other objects* that were part of the task at hand. However, the overall strategy of participants appeared similar to that observed in the baseline, with the routine persisting unchanged. The same applies to the auditory low, auditory high, and mental low conditions, where the routine remained approximately the same as in the baseline condition. In the mental high condition, however, notable differences in behavior were apparent compared to the baseline. Observing these behavior charts, new measures were extracted to inves-

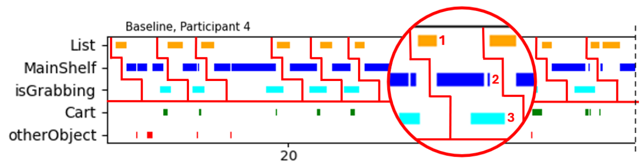
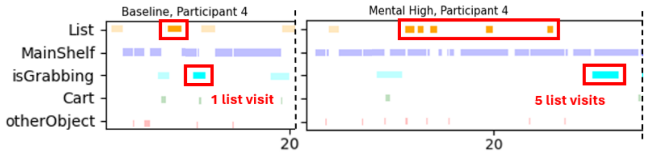
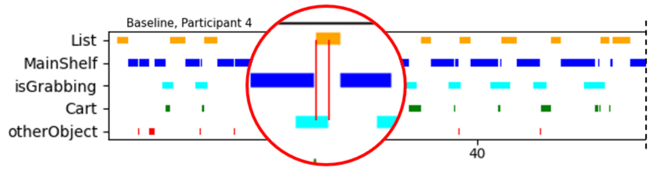


Fig. 10: The behavior chart for the baseline condition for participant four, with recurring routines (as described in III-B) depicted as boxed segments. Initially, participants gaze at the shopping list (1), then locate the product on the main shelf (2), after which they collect the item (3). After these steps the routine repeats.



(a) The average number of gaze visits to the shopping list per collected item. On the left, the red boxes illustrate that the list is visited once (orange bar) for one collected item (cyan bar), while on the right the list is visited five times.



(b) Visual representation of the gazing at shopping list while grabbing measure. The red lines illustrate the overlap of looking at the list while grabbing an object.

Fig. 11: Visual representation of the measures; (a) mean shopping list visits and (b) gazing at shopping list while grabbing measure.

tigate differences in behavior compared to baseline:

- *Mean shopping list visits*: The average number of gaze visits to the shopping list per collected item was computed (Figure 11a). Apart from a drastic drop in the performance of product collection, participants seem to revisit the list more frequently per item that they collected. This may indicate that participants forget what item they were collecting and revisit the list to check again.
- *Mean shopping list time*: The average time gazing at the shopping list per collected item was computed. Participants seem to spend more time on the list per item that they collected compared to the baseline condition. Some participants seem to gaze more at the list for a longer time without revisiting while some spend more time at the list by navigating back multiple times per item collected.
- *Gazing at shopping list while grabbing*: The percentage of time allocated to gazing at the shopping list while collecting an object was used as a measure (Figure 11b). Despite being instructed that the task should not be approached as a race, participants appear to cease consulting the list while collecting the item to save time.

Table IV presents the results of the LME model with Holm's p-value correction and the effect sizes (Cohen's d), with significant p-values emphasized in bold text. Firstly, in comparison to the baseline, the average number of gaze visits to the shopping list per collected item significantly increased in the mental high condition ($d = 2.98$, $p < 0.001$). Secondly, the average time spent gazing at the shopping list per collected item was significantly higher ($d = 3.23$, $p < 0.001$) in the mental high condition compared to the baseline. Lastly, it was observed that the percentage of time allocated to gazing at the *shopping list* while grabbing an object in the mental high condition, in contrast to the baseline, exhibited a medium effect, yet a significant difference ($d = -0.59$, $p = 0.038$). Conversely, the visual low and high, auditory low and high, and the mental low conditions showed no significant differences compared to the baseline.

IV. Discussion

A. Impact of the Different Stimuli

In this study, 22 participants performed a shopping task in an immersive virtual environment using an HMD in seven different conditions: a baseline without extra stimuli, a low and high-level visual condition with additional characters and events, a low and high-level auditory condition with additional background noise and a low and high-level mental condition, including the 0-back task and the 2-back task. The objective of this research was to investigate the impact of different levels of visual, auditory, and mental stimuli on cognitive load. Measures linked to cognitive load and underexplored measures were used to do the analysis. The analysis used a linear mixed model to find differences between the other conditions relative to the baseline condition. On top of that, a combination of AOIs, performance, and grabbing interaction was used for an exploratory behavior analysis, to analyze how the different stimuli affected the behavior and tactics of the participants.

The picture that emerged from this research is that the mental stimuli had the largest effect on cognitive load. Where the mental low condition shows significant differences compared to the baseline for the gaze duration on the *shopping list* and the weighted NASA-TLX score, the mental high condition shows larger effects and significant differences compared to the baseline on performance, NASA-TLX score, head stillness duration and gaze duration on the *main shelf* and *shopping list*. The impact of visual and auditory stimuli, however, barely showed any differences compared to the baseline. The visual condition only showed a difference compared to the baseline for the gaze duration on *other objects*, and in the auditory conditions, only the auditory high condition showed a significant difference compared to the baseline for the subjective NASA-TLX score.

Similar to the paper of Cucinella et al. [18], the findings of the impact of different types of stimuli align with the multiple resource theory [45]. According to this theory, optimal performance is achieved when tasks involve distinct resource channels, thereby preventing interference and minimizing the chances of errors. Conversely, when tasks compete for the same resource channel, performance may be compromised due to an overload of resources. In this research, the main task and the n-back task were both cognitively demanding since they both required information processing and memorization, which may be the reason for the impact of the mental stimuli.

Comparing the different levels of stimuli, also some differences emerge. The mental high condition shows significant differences compared to the baseline for the gaze duration on the *main shelf*, the head stillness duration, and the performance, while the mental low condition does not show significant differences compared to the baseline for these measures. This finding is consistent with prior research, which has shown that varying the difficulty levels of the n-back task results in varying levels of cognitive load [35]. Also, some

TABLE IV: Results from the linear mixed effects model. The reported effect size Cohen's d is highlighted in bold text if the corrected p-value was significant ($p < 0.05$). The shopping list is an area of interest as described in II-E. The description of the measures can be found in III-B. [1]Grabbing refers to the moment when participants grab a product from the main shelf until releasing the product in the cart.

Measure	Visual Low	Visual High	Auditory Low	Auditory High	Mental Low	Mental High	Unit
Mean *shopping list* visits	-0.25	0.20	0.36	0.14	0.81	**2.98**	–
Mean *shopping list* time	-0.02	-0.02	0.41	0.36	0.78	**3.23**	*seconds*
Gazing at *shopping list* while grabbing[1]	-0.15	-0.16	-0.07	0.13	0.53	**-0.59**	%

differences are visible in the visual conditions, where the level of distraction (gaze duration on *other objects*) has a larger effect in the visual high condition than the visual low condition compared to baseline. Despite the pronounced differences in distraction levels between the low and high visual conditions (with higher distraction noted in the high condition), the responses of other measures seem similar. For example, both the NASA-TLX and performance, typically linked with cognitive load, demonstrate variations between the low and high visual conditions compared to the baseline. Nevertheless, these differences are too subtle to state clear distinctions. The same applies to the auditory conditions, where except for the NASA-TLX score, the effects compared to the baseline for all other measures appear similar. The similarity between the low and high conditions for the visual and auditory stimuli makes it difficult to compare them and derive conclusions from the impact of factors that may influence cognitive load in these conditions. One could question if the various levels of stimuli do not influence the impact on cognitive load at all, or if their effect is simply not apparent until the quantity of visual and auditory stimuli is increased.

B. Response of the Measures to Cognitive Load

In addition to revealing insights into visual, auditory, and mental demands, the findings provide insights into the usability of various measures for assessing cognitive load. Results for the performance yield the same result in the research of Cucinella et al. [18], where the largest effect was visible for the mental task. The NASA-TLX scores are almost in line with the research of [18], where the largest effects are visible for the mental condition. Interestingly, the subjective NASA-TLX scores are higher relative to baseline in the auditory high condition, while the performance and psychophysiological responses for this condition show small effects compared to baseline. On the one hand, it can be that cognitive load is affected by the auditory stimuli, but that performance and psychophysiological responses are not sensitive to it. On the other hand, the NASA-TLX has been criticized in that it may not always accurately reflect how much mental effort a person is really using [46]. Moreover, auditory stimuli may affect other factors such as stress, which is linked to cognitive load but not the same [47], participants potentially confusing it with cognitive load and misestimating it subjectively.

Even though heart rate has been shown to increase for higher levels of cognitive load [29], [30], our research does not show significant differences. However, when only the first 35 seconds were considered, the heart rate was significantly higher in both mental conditions. There may be multiple contributing explanations for this. First of all, this may happen because in these conditions participants were required to say aloud the digits, which may increase physical activity and breathing patterns, which in turn affect the heart rate [48]. Furthermore, the heart has a higher sensitivity to more pronounced changes in stimuli, as suggested by Turpin and Siddle [49]. A substantial response may be triggered by the substantial change in stimulus intensity by the n-back task. Moreover, Clifton et al. [50] demonstrated that exposure to a constant stimulus induces a peak heart rate response, followed by a habituation period during which the heart rate decreases again. In this research, the participant might experience decreased arousal as the condition progresses due to the task's relative constancy throughout the session.

In addition to the measures used in the literature, under-explored measures such as gaze duration on different areas of interest, hand movement smoothness, and head stillness duration were investigated to see if they provide insights to assess cognitive load. The hand movement smoothness did not show large effects or significant differences compared to the baseline. Unlike performance and NASA-TLX, known to be indicators for cognitive load, hand trajectory smoothness remained consistent across all stimuli conditions, suggesting its limited suitability for assessing cognitive load. One potential reason could be that there is just no difference in smoothness for the hand trajectories. Despite an increase in cognitive load, the movement involved in selecting a product may remain unchanged, potentially resulting in no discernible difference compared to the baseline. Additionally, even if there were disparities in hand smoothness trajectory, they might not be directly apparent due to variations in distance between each product and the cart, with the average value possibly hiding changes in smoothness.

The head stillness is significantly higher in the mental high condition and shows no large effects nor significant differences in the other conditions compared to baseline. Also, head stillness duration is negatively correlated to performance ($r = -0.62$), and performance is often a reliable indicator of cognitive load. Thus, the findings suggest a potential relationship between head stillness duration and cognitive load, supporting the hypothesis that participants concentrate on a fixed point to minimize head movements, similar to the decrease in fixation rate with higher cognitive load as observed in previous studies [32], [33]. However, this does not imply that head stillness duration can generally serve as an indicator of cognitive load. While participants cannot control psychophysiological responses, head movement is task-dependent; for example, in tasks requiring screen focus, head movements are naturally restricted and may not be suitable for assessing cognitive load. Therefore, while the measure showed promising results for its relation with cognitive load, further investigation is necessary to determine its broader applicability.

Finally, multiple differences are visible in the gaze duration focused on specific AOIs. In essence, these measures are best analyzed collectively as the distribution of gaze duration spent on AOIs, as they are inherently interconnected; the cumulative time spent on the four AOIs adds up to the entirety of one condition, thus affecting the distribution of time to an AOI. This also explains why the correlation between gaze duration on the *main shelf* and *shopping list* are negatively correlated ($r = -0.72$); when participants gaze more at the main shelf, they have less time left to gaze at the list, and the other way around. However, the distribution of the gaze time is dependent on the type of stimulus. Participants spent more gaze time at the *shopping list* in the mental high condition, potentially because they forgot what product they were looking for while processing the n-back task. In the visual conditions, on the other hand, participants spent more time at *other objects* because of the visual distractions. When comparing the gaze pattern to the NASA-TLX and performance, the results suggest that the visual distraction (gaze duration on *other objects*) does not impact the cognitive load, while the gaze focus on the *shopping list* can be linked to cognitive load. Gaze behavior may therefore serve as a valuable tool for analyzing cognitive load, but these results show that a change in gaze behavior does not directly indicate an increase in cognitive load. Moreover, its applicability is not generalizable and should be adjusted to the specific task being assessed.

C. Exploratory Behavior Analysis

Besides the analysis of the measures, this research performed an exploratory behavior analysis. For the behavior analysis, gaze behavior, performance, and grabbing interaction were combined to see if this could be used to analyze the impact of the different stimuli, and ultimately on the cognitive load. As far as concerned, no research has used this method to analyze the cognitive load. The analysis showed that participants used a routine to execute the task at hand. Three measures were extracted, which all showed significant differences compared to the baseline in the mental high condition. Especially the features linked to the routine (mean *shopping list* visits and time) showed large effects and a significant difference compared to the baseline in the mental high condition. These results support the finding that mental stimuli had the largest effect on cognitive load, while the other stimuli showed little effect. Moreover, the behavior analysis provides an intuitive method to see how the measures affect the cognitive load, and how the increase of cognitive load ultimately translates to the actions of the participant.

For example, by performing the analysis it is possible to understand whether the visual stimuli affect the behavior of the participants. Examining the behavior chart in the visual high condition, it was visible that participants gazed for a longer time at *other objects* compared to the baseline condition. However, the routine in the visual high condition did not change compared to the baseline, which was also determined statistically. So, even though the participant was more distracted, it is likely to say that this does not have a significant impact on the behavior. This supports that visual stimuli have a limited impact on the cognitive load compared to mental stimuli. Furthermore, the analysis showed that the routine is disrupted in the mental high condition. Besides supporting the fact that mental stimuli impact the cognitive load more compared to visual and auditory stimuli, it also gives insights into how the increased cognitive load ultimately affects the actions of the participant. In this case for example, when cognitive load increased, the participants appeared to fail to memorize the item that they were looking for, resulting in revisiting the list.

Of course, the measures derived from the behavior charts are influenced by performance. Specifically, when the number of products collected is very low, the number of visits to the *shopping list* tends to be higher. However, it is important to note that if participants exhibited the same behavior as in the baseline, the mean *shopping list* visits and the mean *shopping list* time would not be concurrently affected. The results demonstrate an increase in both measures, indicating a disruption in the routine as cognitive load intensifies. This disruption, in turn, impacts the performance of the participant. Thus, while there is a relation between behavior and performance, it is clear that they provide different and valuable information.

D. Limitations of the Study

This study also has some limitations. First, some material limitations should be considered when interpreting the results. First of all, gaze data from the HMD was presented in 3-dimensional arrays rounded to two decimal places. With an interval of around 0.011 seconds between frames, minor movements could yield disproportionately high differences. This rounding-off sometimes made it look like participants were oscillating when, in reality, they were stationary. Higher data accuracy for the head-mounted display would improve overall precision. Secondly, the sensor error in the tracking device of the HMD may have impacted the calculations of the movement measures. This, in combination with computations involving functions such as the third derivative for the jerk, can have a substantial impact on the outcome. Thirdly, the bounding box of the controller, designed to register grabbable items, has a trade-off in size. A bounding box that was too large led to inaccurate grabbing, while a too-small one made grabbing difficult. This resulted in participants occasionally grabbing unintended items, particularly when trying to grab smaller ones, impacting both accuracy and the average number of items collected per minute.

Furthermore, in addition to material limitations, it is important to note that the sample size in this experiment was relatively small ($n = 22$). The limited number of participants causes variability within the group of samples, making it challenging to distinguish true effects from random fluctuations. The reduced sample size also decreases the power of the statistical test, making it more sensible to outliers in the small group of participants. Moreover, the small sample size made it impossible to counterbalance potential condition combinations, making it difficult to determine the impact of the condition order on the results.

Another limitation of the study is the possibility that the time interval between consecutive conditions may not have been adequate for allowing psychophysiological responses from the previous condition to fully return to baseline levels. One notable example is the impact on heart rate recovery after each condition. Unfortunately, the study did not verify whether the heart rate had returned to each participant's baseline before starting the next condition. This oversight raises the possibility that the previous condition may have affected the subsequent heart rate measurements.

E. Future Work

This research primarily focused on comparing various levels of stimuli to a baseline. However, it could be valuable to explore the combined impact of different stimuli or compare different conditions directly. For instance, analyzing the levels of the same stimulus may reveal underlying differences not apparent when considering each level individually against the baseline. Additionally, combining stimuli may affect cognitive load differently than when stimuli are presented separately. Further investigation into these comparisons and stimuli combinations could provide insights into how these stimuli ultimately impact the cognitive load.

Furthermore, research has been conducted with healthy participants, while individuals with cognitive deficits may elicit different responses. For example, it has been shown that psychophysiological responses in neurorehabilitation patients show similarities, albeit weaker, to those observed in healthy individuals. Furthermore, these responses can vary among patients depending on factors such as the type of impairment [51] or age [52]. Consequently, it is unlikely that the results of this research can be generalized for everyone, yet it gives a direction. Future research could focus on the impact of these stimuli on different subject groups which may elicit different responses, such as individuals with cognitive deficits.

Lastly, future work could focus on developing an algorithm capable of real-time classification of cognitive load using various measures. The effectiveness of rehabilitation involves making sure that patients do not experience too much cognitive load or be under-challenged. By creating the algorithm, it is possible to monitor the cognitive load in real-time of patients undergoing neurorehabilitation and adjust the environment to the needs of the patient. Knowing which stimuli significantly impact the cognitive load can help in making design decisions about which stimuli to add when the cognitive load is too low, and which ones to remove when the cognitive load is too high.

V. Conclusion

This research investigated the impact of different levels of visual, auditory, and mental stimuli on various measures and cognitive load in a virtual shopping task. The findings revealed that mental stimuli had the most significant impact on cognitive load, while visual stimuli were more distracting but did not significantly affect cognitive load. Auditory stimuli showed minimal differences compared to baseline, indicating limited influence on cognitive load. Additionally, differences were observed between low and high conditions, particularly evident in the mental stimuli, where the high condition increased cognitive load more than the low condition. Furthermore, a range of measures, including both established and underexplored ones, were employed to assess cognitive load and the impact of stimuli. Performance and NASA-TLX scores aligned as expected, while heart rate and heart rate variability showed no significant effects. New measures such as gaze behavior and head stillness duration were linked to cognitive load, whereas hand movement smoothness did not exhibit significant differences across conditions. Finally, the results regarding the impact of the stimuli were corroborated by behavior analysis, which integrated gaze behavior, performance, and grabbing interaction, which offered an intuitive method to analyze the behavior and actions of the participants. This analysis highlighted substantial differences in routine under mental stimuli, while no significant differences were observed for visual and auditory stimuli. Future research is thus advised to consider the finding that mental stimuli affect cognitive load more than visual and auditory stimuli when designing an immersive virtual environment. Finally, it is recommended to explore the impact of these stimuli on different types of subject groups, such as patients with cognitive deficits, since the responses may differ. Such research could be a step towards dynamically adjusting the immersive virtual environment in rehabilitation by monitoring the cognitive load in real-time.

References

[1] M. Oberholzer and R. M. Müri, "Neurorehabilitation of traumatic brain injury (tbi): A clinical review," *Medical Sciences*, vol. 7, no. 3, p. 47, 2019.

[2] P. Langhorne, J. Bernhardt, and G. Kwakkel, "Stroke rehabilitation," *The Lancet*, vol. 377, no. 9778, pp. 1693–1702, 2011.

[3] N. A. Bayona, J. Bitensky, K. Salter, and R. Teasell, "The role of task-specific training in rehabilitation therapies," *Topics in stroke rehabilitation*, vol. 12, no. 3, pp. 58–65, 2005.

[4] C. Bütefisch, H. Hummelsheim, P. Denzler, and K.-H. Mauritz, "Repetitive training of isolated movements improves the outcome of motor rehabilitation of the centrally paretic hand," *Journal of the neurological sciences*, vol. 130, no. 1, pp. 59–68, 1995.

[5] G. Nelles, "Cortical reorganization–effects of intensive therapy," *Restorative neurology and neuroscience*, vol. 22, no. 3-5, pp. 239–244, 2004.

[6] W. R. Sherman and A. B. Craig, "Understanding virtual reality," *San Francisco, CA: Morgan Kauffman*, 2003.

[7] E. A. Keshner and A. Lamontagne, "The untapped potential of virtual reality in rehabilitation of balance and gait in neurological disorders," *Frontiers in virtual reality*, vol. 2, p. 641 650, 2021.

[8] G. Gonçalves, H. Coelho, P. Monteiro, M. Melo, and M. Bessa, “Systematic review of comparative studies of the impact of realism in immersive virtual experiences,” *ACM Computing Surveys*, vol. 55, no. 6, pp. 1–36, 2022.

[9] M. Newman, B. Gatersleben, K. Wyles, and E. Ratcliffe, “The use of virtual reality in environment experiences and the importance of realism,” *Journal of environmental psychology*, vol. 79, p. 101 733, 2022.

[10] M. Slater, P. Khanna, J. Mortensen, and I. Yu, “Visual realism enhances realistic response in an immersive virtual environment,” *IEEE computer graphics and applications*, vol. 29, no. 3, pp. 76–84, 2009.

[11] M. I. Berkman and E. Akan, “Presence and immersion in virtual reality,” in *Encyclopedia of Computer Graphics and Games*, N. Lee, Ed. Cham: Springer International Publishing, 2019, pp. 1–10, ISBN: 978-3-319-08234-9. DOI: 10.1007/978-3-319-08234-9_162-1. [Online]. Available: https://doi.org/10.1007/978-3-319-08234-9_162-1.

[12] G. Vecchiato, G. Tieri, A. Jelic, F. De Matteis, A. G. Maglione, and F. Babiloni, “Electroencephalographic correlates of sensorimotor integration and embodiment during the appreciation of virtual architectural environments,” *Frontiers in psychology*, vol. 6, p. 1944, 2015.

[13] M. Hallett, “Neuroplasticity and rehabilitation,” *Journal of Rehabilitation Research and Development*, vol. 42, no. 4, R17, 2005.

[14] V. L. Roger, A. S. Go, D. M. Lloyd-Jones, *et al.*, “Heart disease and stroke statistics—2011 update: A report from the american heart association,” *Circulation*, vol. 123, no. 4, e18–e209, 2011.

[15] C. D. Wickens, “Workload assessment and prediction,” *MANPRINT: an approach to systems integration*, p. 257, 2012.

[16] G. Orru and L. Longo, “The evolution of cognitive load theory and the measurement of its intrinsic, extraneous and germane loads: A review,” in *Human Mental Workload: Models and Applications: Second International Symposium, H-WORKLOAD 2018, Amsterdam, The Netherlands, September 20-21, 2018, Revised Selected Papers 2*, Springer, 2019, pp. 23–48.

[17] J. Sweller, “Cognitive load theory, learning difficulty, and instructional design,” *Learning and instruction*, vol. 4, no. 4, pp. 295–312, 1994.

[18] S. L. Cucinella, J. C. de Winter, A. van den Berg, *et al.*, “Shopping in immersive virtual reality: Effects of diminishing visual, auditory, and cognitive demands on workload,”

[19] C. Colombatto, B. Van Buren, and B. J. Scholl, “Intentionally distracting: Working memory is disrupted by the perception of other agents attending to you—even without eye-gaze cues,” *Psychonomic bulletin & review*, vol. 26, pp. 951–957, 2019.

[20] B. Arons, “A review of the cocktail party effect,” *Journal of the American Voice I/O society*, vol. 12, no. 7, pp. 35–50, 1992.

[21] G. K. Kountouriotis, R. M. Wilkie, P. H. Gardner, and N. Merat, “Looking and thinking when driving: The impact of gaze and cognitive load on steering,” *Transportation research part F: traffic psychology and behaviour*, vol. 34, pp. 108–121, 2015.

[22] N. Assink, R. Lubbe, J.-P. Fox, Y. Wang, B. Pierre, and I. Rudas, “Does time pressure induce tunnel vision? an examination with the eriksen flanker task by applying the hierarchical drift diffusion model,” in *Proceedings of the International Conference on Neural Networks–Fuzzy Systems (NN-FS 2015)*, 2015, pp. 30–40.

[23] F. G. Paas, J. J. Van Merriënboer, and J. J. Adam, “Measurement of cognitive load in instructional research,” *Perceptual and motor skills*, vol. 79, no. 1, pp. 419–430, 1994.

[24] S. G. Hart and L. E. Staveland, “Development of nasa-tlx (task load index): Results of empirical and theoretical research,” in *Advances in psychology*, vol. 52, Elsevier, 1988, pp. 139–183.

[25] T. Luong, N. Martin, A. Raison, F. Argelaguet, J.-M. Diverrez, and A. Lécuyer, “Towards real-time recognition of users mental workload using integrated physiological sensors into a vr hmd,” in *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, IEEE, 2020, pp. 425–437.

[26] Y. Shi, N. Ruiz, R. Taib, E. Choi, and F. Chen, “Galvanic skin response (gsr) as an index of cognitive load,” in *CHI’07 extended abstracts on Human factors in computing systems*, 2007, pp. 2651–2656.

[27] S. Siriya, M. Lochner, A. Duenser, and R. Taib, “Exploring novel methodology for classifying cognitive workload,” in *Intelligent Technologies for Interactive Entertainment: 10th EAI International Conference, INTETAIN 2018, Guimarães, Portugal, November 21-23, 2018, Proceedings 10*, Springer, 2019, pp. 105–114.

[28] B. Mehler, B. Reimer, J. F. Coughlin, and J. A. Dusek, “Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers,” *Transportation research record*, vol. 2138, no. 1, pp. 6–12, 2009.

[29] D. Novak, J. Ziherl, A. Olenšek, *et al.*, “Psychophysiological responses to robotic rehabilitation tasks in stroke,” *IEEE Transactions on neural systems and rehabilitation engineering*, vol. 18, no. 4, pp. 351–361, 2010.

[30] N. Goljar, M. Javh, J. Poje, *et al.*, “Psychophysiological responses to robot training in different recovery phases after stroke,” in *2011 IEEE International Conference on Rehabilitation Robotics*, IEEE, 2011, pp. 1–6.

[31] J. Ryu and E. B. Torres, “Characterization of sensory-motor behavior under cognitive load using a new statistical platform for studies of embodied cognition,” *Frontiers in human neuroscience*, vol. 12, p. 116, 2018.

[32] S. Das, J. Maiti, and O. Krishna, “Assessing mental workload in virtual reality based eot crane operations: A multi-measure approach,” *International Journal of Industrial Ergonomics*, vol. 80, p. 103 017, 2020.

[33] G. Marquart, C. Cabrall, and J. de Winter, “Review of eye-related measures of drivers’ mental workload,” *Procedia Manufacturing*, vol. 3, pp. 2854–2861, 2015.

[34] K. Walter and P. Bex, “Cognitive load influences oculomotor behavior in natural scenes,” *Scientific Reports*, vol. 11, no. 1, p. 12 405, 2021.

[35] S. M. Jaeggi, M. Buschkuehl, W. J. Perrig, and B. Meier, “The concurrent validity of the n-back task as a working memory measure,” *Memory*, vol. 18, no. 4, pp. 394–412, 2010.

[36] F. Paas, J. E. Tuovinen, H. Tabbers, and P. W. Van Gerven, “Cognitive load measurement as a means to advance cognitive load theory,” *Educational psychologist*, vol. 38, no. 1, pp. 63–71, 2003.

[37] C. Carreiras, A. P. Alves, A. Lourenço, F. Canento, H. Silva, A. Fred, *et al.*, *BioSPPy: Biosignal processing in Python*, [Online; accessed ¡today¿], 2015–. [Online]. Available: https://github.com/PIA-Group/BioSPPy/.

[38] T. Flash and N. Hogan, “The coordination of arm movements: An experimentally confirmed mathematical model,” *Journal of neuroscience*, vol. 5, no. 7, pp. 1688–1703, 1985.

[39] C. S. K. Dash, A. K. Behera, S. Dehuri, and A. Ghosh, “An outliers detection and elimination framework in classification task of data mining,” *Decision Analytics Journal*, vol. 6, p. 100 164, 2023.

[40] G. Barbato, E. Barini, G. Genta, and R. Levi, “Features and performance of some outlier detection methods,” *Journal of Applied Statistics*, vol. 38, no. 10, pp. 2133–2149, 2011.

[41] A. Gałecki, T. Burzykowski, A. Gałecki, and T. Burzykowski, *Linear mixed-effects model*. Springer, 2013.

[42] S. Holm, “A simple sequentially rejective multiple test procedure,” *Scandinavian Journal of Statistics*, vol. 6, pp. 65–70, 1979. [Online]. Available: https://www.jstor.org/stable/4615733.

[43] J. Cohen, *Statistical power analysis for the behavioral sciences*. Academic press, 2013.

[44] F. Faul, E. Erdfelder, A.-G. Lang, and A. Buchner, “G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences,” *Behavior research methods*, vol. 39, no. 2, pp. 175–191, 2007.

[45] C. D. Wickens, “Multiple resources and performance prediction,” *Theoretical issues in ergonomics science*, vol. 3, no. 2, pp. 159–177, 2002.

[46] R. D. McKendrick and E. Cherry, “A deeper look at the nasa tlx and where it falls short,” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, SAGE Publications Sage CA: Los Angeles, CA, vol. 62, 2018, pp. 44–48.

[47] A. Woody, E. D. Hooker, P. M. Zoccola, and S. S. Dickerson, “Social-evaluative threat, cognitive load, and the cortisol and cardiovascular stress response,” *Psychoneuroendocrinology*, vol. 97, pp. 149–155, 2018.

[48] P. Liehr, “Uncovering a hidden language: The effects of listening and talking on blood pressure and heart rate,” *Archives of Psychiatric Nursing*, vol. 6, no. 5, pp. 306–311, 1992.

[49] G. Turpin and D. A. Siddle, “Effects of stimulus intensity on cardiovascular activity,” *Psychophysiology*, vol. 20, no. 6, pp. 611–624, 1983.

[50] R. K. Clifton, F. K. Graham, and H. M. Hatton, “Newborn heart-rate response and response habituation as a function of stimulus duration,” *Journal of Experimental Child Psychology*, vol. 6, pp. 265–278, 1968.

[51] D. Kumar, A. Dutta, A. Das, and U. Lahiri, “Smarteye: Developing a novel eye tracking system for quantitative assessment of oculomotor abnormalities,” *IEEE Transactions on neural systems and rehabilitation engineering*, vol. 24, no. 10, pp. 1051–1059, 2016.

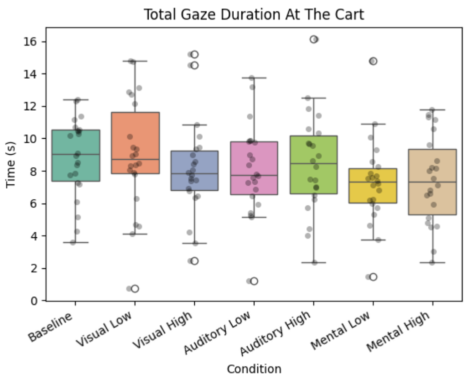
[52] E. Ferreira, D. Ferreira, S. Kim, *et al.*, “Assessing real-time cognitive load based on psycho-physiological measures for younger and older adults,” in *2014 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)*, IEEE, 2014, pp. 39–48.

A

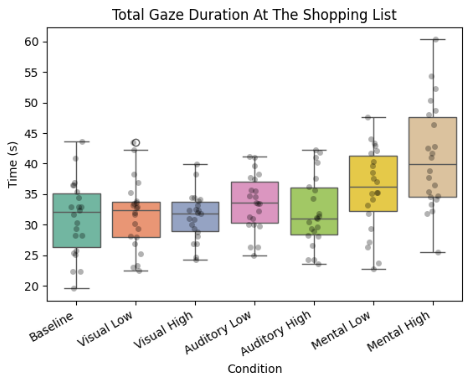
Additional Data & Results

A.1. Box Plots For All Measures

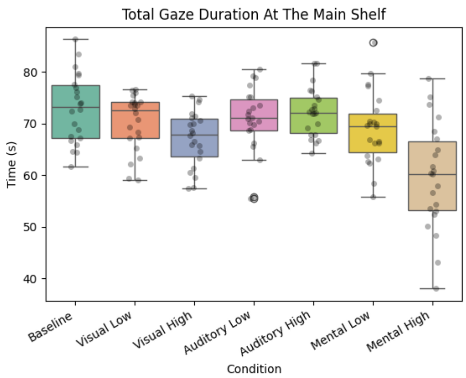
This section contains the box plots for the dependent measures from the paper in Figures A.1 and A.2. The box plots for the behavior chart are added at the end of the section in Figure A.3.



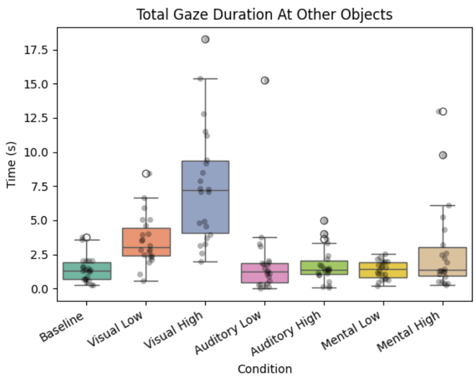
(a) Box plot showing the total gaze duration on the *Cart* across conditions.



(b) Box plot showing the total gaze duration on the *Shopping List* across conditions.

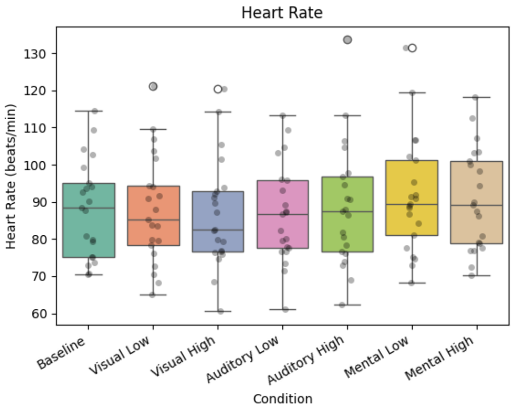


(c) Box plot showing the total gaze duration on the *Main Shelf* across conditions.

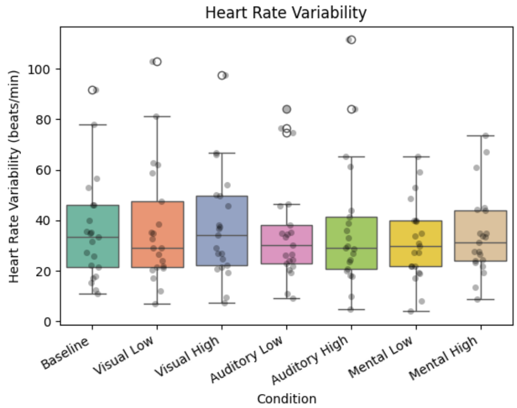


(d) Box plot showing the total gaze duration on the *Other Objects* across conditions.

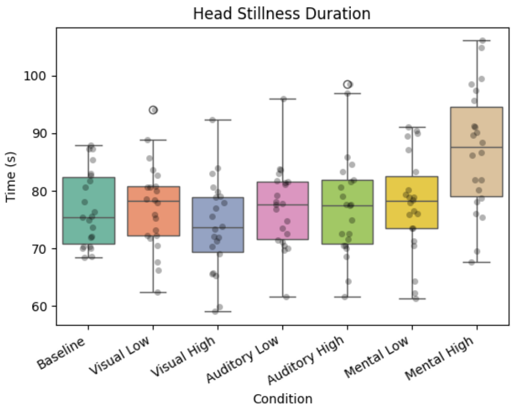
Figure A.1: Box plots showing the total gaze duration on specific areas of interest during each condition with participant values represented by a jitter plot.



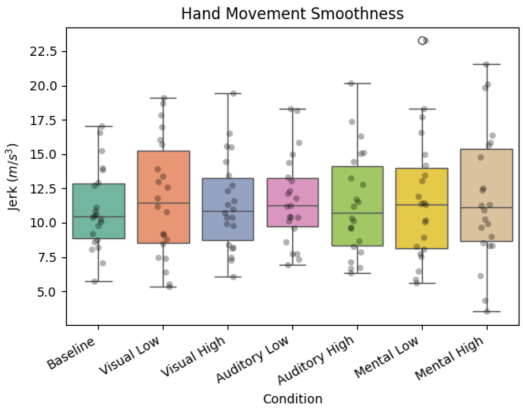
(a) Box plot with the mean heart rate of all participants across conditions.



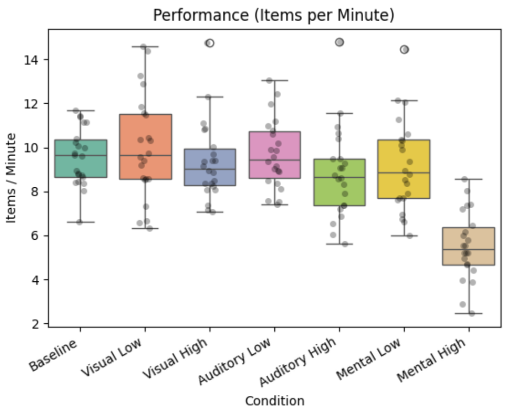
(b) Box plot with the mean heart rate variability of all participants across conditions.



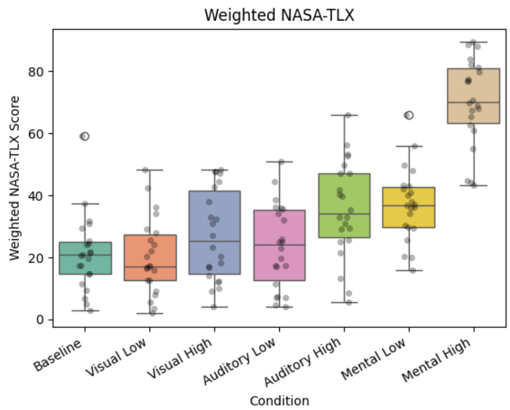
(c) Box plot with the head stillness duration, i.e. the total time the head acceleration is below 3 m/s^2, across conditions.



(d) Box plot with the hand movement smoothness, i.e. the mean jerk of all trajectories where a participant was holding a product, across conditions.



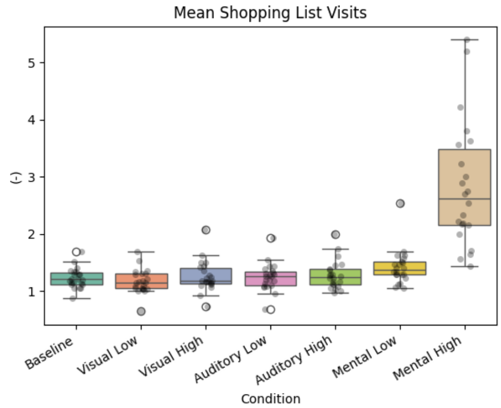
(e) Box plot with the performance, i.e. the number of items collected per minute, across conditions.



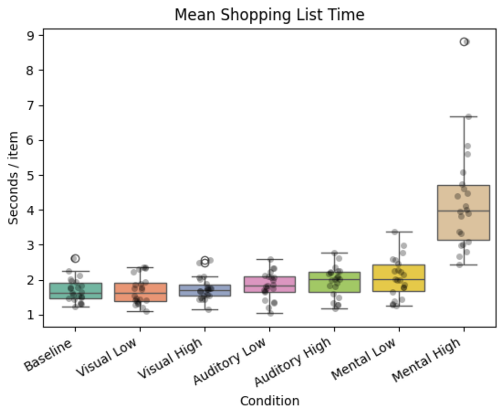
(f) The weighted NASA-TLX score on a scale from 0 to 100 across conditions.

Figure A.2: Box plots of heart rate, heart rate variability, hand movement smoothness, head stillness duration, performance, and subjective NASA-TLX score during each condition with participant values represented by a jitter plot.

(a) Box plot with the percentage of time allocated to gazing at the *shopping list* while collecting an object across conditions.



(b) Box plot with the average number of gaze visits to the *shopping list* per collected item across conditions.



(c) Box plot with the average time gazing at the *shopping list* per collected item across conditions.

Figure A.3: Box plots of the behavior features: the percentage of time allocated to gazing at the *shopping list* while collecting an object, the average number of gaze visits to the *shopping list* per collected item, and the average time gazing at the *shopping list* per collected item.

A.2. Feature Overview Per Condition

This section provides an overview in the format of tables with the means, standard deviation, skewness, and kurtosis for each condition (Table A.1). Each table represents one condition, due to space constraints. Literature is not always unanimous when speaking about ”normal” values for skewness and kurtosis. George and Mallery [1] considered the values for asymmetry and kurtosis between -2 and +2 acceptable to prove normal univariate distribution, while Hair et al. [2] argued that data is considered to be normal if skewness is between -2 to +2 and kurtosis is between -7 to +7. This research uses the most used boundaries, which are between -2 and +2 and between -7 and +7.

After inspecting the data, there are no conditions where the limits of the skewness or kurtosis are exceeded. Even though the data is not normally distributed for every condition according to the Shapiro-Wilk test, the assumption was made that all data was distributed in a normal way. All plots and calculations could be reviewed using the code in this Github repository

Multidimensional Analysis of Cognitive, Physiological, and Performance Metrics in Various Conditions

Condition	Mean	Min.	Max.	SD	Skewness	Kurtosis	Unit
Gaze Duration on *Cart*							
Baseline	8.819	2.431	3.574	12.401	-0.547	-0.565	*seconds*
Visual Low	8.937	3.485	0.727	14.791	-0.244	-0.215	*seconds*
Visual High	8.107	2.914	2.468	15.236	0.584	0.836	*seconds*
Auditory Low	8.077	2.731	1.216	13.777	-0.065	0.536	*seconds*
Auditory High	8.440	3.047	2.345	16.127	0.300	0.243	*seconds*
Mental Low	7.291	2.609	1.460	14.825	0.587	1.783	*seconds*
Mental High	7.444	2.701	2.356	11.781	-0.004	-0.921	*seconds*
Gaze Duration on *Shopping List*							
Baseline	31.076	5.990	19.566	43.636	0.030	-0.532	*seconds*
Visual Low	31.657	5.572	22.421	43.556	0.248	-0.364	*seconds*
Visual High	31.354	3.827	24.207	39.960	0.134	-0.075	*seconds*
Auditory Low	33.549	4.513	24.938	41.138	-0.104	-0.761	*seconds*
Auditory High	32.161	5.620	23.633	42.226	0.370	-0.912	*seconds*
Mental Low	35.870	6.683	22.781	47.576	-0.365	-0.742	*seconds*
Mental High	41.060	8.437	25.526	60.339	0.440	-0.472	*seconds*
Gaze Duration on *Main Shelf*							
Baseline	72.988	6.548	61.686	86.326	0.180	-0.882	*seconds*
Visual Low	70.233	5.360	59.048	76.589	-0.753	-0.633	*seconds*
Visual High	67.122	5.236	57.522	75.388	-0.315	-0.900	*seconds*
Auditory Low	70.599	6.431	55.433	80.576	-0.795	0.425	*seconds*
Auditory High	72.289	4.668	64.228	81.623	0.327	-0.551	*seconds*
Mental Low	69.062	6.875	55.894	85.751	0.389	0.042	*seconds*
Mental High	59.753	10.127	38.088	78.796	-0.086	-0.479	*seconds*
Gaze Duration on *Other Objects*							
Baseline	1.414	0.907	0.249	3.773	1.105	0.984	*seconds*
Visual Low	3.522	1.807	0.568	8.415	0.884	0.591	*seconds*
Visual High	7.535	4.179	1.975	18.247	0.844	0.121	*seconds*
Auditory Low	1.372	1.035	0.000	3.783	0.659	-0.259	*seconds*
Auditory High	1.708	1.250	0.057	5.001	1.063	0.516	*seconds*
Mental Low	1.379	0.637	0.197	2.511	-0.087	-1.095	*seconds*
Mental High	2.787	3.167	0.234	13.010	1.962	3.204	*seconds*

Table A.1 – Continued from previous page

Condition	Mean	SD	Min.	Max.	Skewness	Kurtosis	Unit
Heart Rate							
Baseline	88.097	12.563	70.479	114.440	0.324	-0.856	bpm
Visual Low	88.112	14.095	65.028	121.068	0.421	-0.404	bpm
Visual High	87.047	14.088	60.585	120.340	0.563	0.036	bpm
Auditory Low	87.172	12.834	61.094	113.415	0.239	-0.503	bpm
Auditory High	88.620	15.771	62.421	133.613	0.918	1.001	bpm
Mental Low	92.048	15.010	68.344	131.420	0.736	0.368	bpm
Mental High	90.777	13.164	70.295	118.163	0.294	-0.910	bpm
Heart Rate Variability							
Baseline	35.580	19.883	10.842	91.627	1.240	1.222	bpm
Visual Low	36.601	23.161	7.216	102.990	1.298	1.117	bpm
Visual High	36.483	20.828	7.304	97.561	1.118	1.186	bpm
Auditory Low	35.208	19.545	9.103	84.262	1.245	0.779	bpm
Auditory High	37.082	24.254	4.763	111.421	1.491	2.053	bpm
Mental Low	30.511	15.848	4.089	65.303	0.431	-0.422	bpm
Mental High	34.631	16.007	8.854	73.559	0.858	0.244	bpm
Hand Movement Smoothness							
Baseline	10.984	2.917	5.711	17.054	0.466	-0.455	m/s^3
Visual Low	11.767	4.205	5.37	19.092	0.201	-1.131	m/s^3
Visual High	11.407	3.302	6.07	19.474	0.574	-0.221	m/s^3
Auditory Low	11.642	3.165	6.923	18.304	0.525	-0.418	m/s^3
Auditory High	11.38	3.727	6.354	20.191	0.564	-0.496	m/s^3
Mental Low	11.657	4.41	5.616	23.268	0.778	0.157	m/s^3
Mental High	11.857	4.769	3.54	21.538	0.362	-0.537	m/s^3
Head Stillness Duration							
Baseline	76.902	6.414	68.48	87.91	0.362	-1.24	$seconds$
Visual Low	77.245	7.375	62.39	94.17	0.145	-0.146	$seconds$
Visual High	73.889	7.884	59.05	92.43	0.099	-0.137	$seconds$
Auditory Low	77.13	6.952	61.6	96.07	0.341	0.969	$seconds$
Auditory High	77.363	8.99	61.56	98.61	0.634	0.256	$seconds$
Mental Low	77.751	8.496	61.29	91.11	-0.186	-0.585	$seconds$
Mental High	87.036	10.436	67.65	106.09	0.025	-0.788	$seconds$
Weighted NASA-TLX Score							
Baseline	21.167	11.979	3.000	59.000	1.180	2.294	–
Visual Low	20.318	12.032	2.000	48.333	0.581	-0.300	–
Visual High	27.091	14.377	4.000	48.333	0.156	-1.384	–
Auditory Low	23.621	13.322	4.000	51.000	0.186	-0.961	–
Auditory High	35.636	15.289	5.667	66.000	-0.127	-0.553	–
Mental Low	37.197	11.508	16.000	66.000	0.362	0.244	–
Mental High	70.197	13.773	43.333	89.333	-0.566	-0.578	–

Table A.1 – Continued from previous page

Condition	Mean	SD	Min.	Max.	Skewness	Kurtosis	Unit
Performance							
Baseline	9.559	1.277	6.626	11.674	-0.099	-0.545	$items/minute$
Visual Low	10.006	2.339	6.335	14.576	0.325	-0.69	$items/minute$
Visual High	9.355	1.748	7.096	14.758	1.331	1.959	$items/minute$
Auditory Low	9.692	1.536	7.402	13.031	0.448	-0.575	$items/minute$
Auditory High	8.811	2.017	5.617	14.772	0.941	1.333	$items/minute$
Mental Low	9.199	2.051	5.987	14.455	0.627	-0.016	$items/minute$
Mental High	5.532	1.527	2.466	8.553	0.047	-0.43	$items/minute$
Mean *shopping list* visits							
Baseline	1.229	0.173	0.882	1.692	0.645	0.764	–
Visual Low	1.172	0.207	0.65	1.692	0.218	1.187	–
Visual High	1.262	0.262	0.737	2.071	1.047	2.274	–
Auditory Low	1.251	0.235	0.684	1.933	0.47	2.079	–
Auditory High	1.289	0.245	0.967	2	1.204	1.24	–
Mental Low	1.42	0.303	1.048	2.538	2.087	5.791	–
Mental High	2.834	1.074	1.438	5.4	0.918	0.117	–
Mean *shopping list* time							
Baseline	1.704	0.337	1.22	2.612	0.824	0.345	$seconds$
Visual Low	1.698	0.4	1.091	2.366	0.417	-1.119	$seconds$
Visual High	1.749	0.329	1.143	2.556	0.863	0.621	$seconds$
Auditory Low	1.827	0.386	1.041	2.597	-0.14	-0.541	$seconds$
Auditory High	1.941	0.422	1.18	2.776	-0.189	-0.623	$seconds$
Mental Low	2.064	0.563	1.247	3.367	0.401	-0.451	$seconds$
Mental High	4.262	1.457	2.422	8.813	1.395	2.085	$seconds$
Gazing at *shopping list* while grabbing							
Baseline	24.271	12.062	0	47.622	-0.071	-0.32	%
Visual Low	23.109	13.29	0	46.437	-0.046	-0.847	%
Visual High	22.673	12.553	0	43.483	-0.088	-0.814	%
Auditory Low	24.765	12.65	0	47.008	0.109	-0.724	%
Auditory High	25.649	14.375	0	46.17	-0.2	-1.19	%
Mental Low	31.105	13.478	0	52.841	-0.556	-0.437	%
Mental High	16.988	11.337	0	39.773	0.539	-0.716	%

A.3. NASA-TLX Results

The paper section only discusses the weighted values of the NASA-TLX, while it could also be useful to see individual differences across conditions. Table A.2 gives an overview of the score for each participant across all conditions. Other data about the NASA-TLX scores in this experiment, such as the individual scores for the *Mental demand*, *Physical demand*, *Temporal demand*, *Performance*, *Effort*, and *Frustration* scales and the weights chosen by the participants, could be found in the data repository in the *nasa-tlx-results.xlsx* file. This section displays the weighted and the unweighted scores of the NASA-TLX.

NASA-TLX Scores

Participant No	Baseline	Visual Low	Visual High	Auditory Low	Auditory High	Mental Low	Mental High
Weighted NASA-TLX Scores							
1	25,33	25,67	12,33	25,00	33,00	20,00	77,33
2	6,67	8,00	12,00	7,33	29,00	36,67	84,00
3	21,00	17,00	20,33	17,33	33,00	36,33	88,00
4	14,67	9,00	31,00	17,00	47,00	41,00	89,33
5	19,67	20,33	16,67	23,00	21,33	29,67	65,33
6	31,00	28,00	33,00	34,00	40,33	40,00	67,33
7	31,67	17,33	32,33	26,00	25,67	48,00	44,00
8	11,33	2,00	14,00	7,00	8,67	36,00	44,67
9	37,33	48,33	44,33	51,00	53,33	56,00	69,67
10	5,00	5,67	4,00	7,00	25,00	25,67	76,67
11	21,33	12,67	47,67	19,67	39,67	49,67	68,00
12	29,33	34,00	38,00	36,33	35,33	34,00	61,00
13	21,67	24,00	17,00	35,67	41,67	42,00	55,00
14	59,00	42,33	47,00	44,33	56,33	37,00	70,67
15	9,33	3,67	9,00	4,67	13,33	20,33	43,33
16	24,33	36,33	47,67	32,00	47,00	43,00	76,67
17	20,67	16,33	42,67	11,33	29,33	38,00	82,00
18	3,00	22,00	10,00	4,00	5,67	16,00	88,67
19	17,33	16,67	18,33	17,33	52,67	29,33	81,33
20	14,67	16,00	23,33	25,00	66,00	30,33	62,67
21	17,33	29,00	48,33	38,67	31,00	66,00	69,00
22	24,00	12,67	27,00	36,00	49,67	43,33	79,67
weighted total	**21,17**	**20,32**	**27,09**	**23,62**	**35,64**	**37,20**	**70,20**
Unweighted NASA-TLX Scores							
1	19,17	18,33	11,67	25,83	33,33	22,50	67,50
2	5,00	5,83	10,83	5,83	20,83	26,67	66,67
3	25,00	23,33	24,17	22,50	34,17	35,00	71,67
4	20,83	18,33	39,17	15,83	44,17	40,83	79,17
5	17,50	17,50	16,67	20,00	17,50	25,83	51,67
6	30,83	27,50	30,83	33,33	40,00	39,17	60,00
7	29,17	19,17	29,17	23,33	21,67	44,17	40,83

Table A.2 – Continued from previous page

Participant No	Baseline	Visual Low	Visual High	Auditory Low	Auditory High	Mental Low	Mental High
8	6,67	3,33	15,00	7,50	9,17	29,17	41,67
9	37,50	46,67	44,17	51,67	52,50	56,67	71,67
10	4,17	5,00	3,33	7,50	20,00	18,33	57,50
11	21,67	13,33	45,00	22,50	40,83	45,83	60,83
12	26,67	29,17	33,33	31,67	32,50	28,33	55,00
13	19,17	21,67	16,67	31,67	36,67	35,83	47,50
14	49,17	36,67	44,17	40,00	50,83	34,17	60,00
15	10,00	3,33	7,50	4,17	11,67	19,17	35,00
16	22,50	40,83	37,50	23,33	35,00	45,00	66,67
17	18,33	16,67	32,50	10,83	25,83	33,33	69,17
18	2,50	20,00	9,17	3,33	5,00	14,17	70,83
19	16,67	15,00	19,17	17,50	50,83	28,33	78,33
20	18,33	17,50	19,17	22,50	55,83	24,17	56,67
21	15,00	24,17	40,83	30,00	26,67	55,83	58,33
22	24,17	12,50	24,17	34,17	44,17	42,50	69,17
unweighted total	**20,00**	**19,81**	**25,19**	**22,05**	**32,23**	**33,86**	**60,72**

A.4. Condition Order

The order of the conditions has been added to the statistical model to incorporate learning effects. The order of the conditions has been randomized using a with randomization tool (http://random.org) since the counterbalancing method was not possible with 22 participants (you would need 5040 participants for perfect counterbalancing). Table A.3 illustrates the ordering of conditions for each participant, also available in the data repository within the *participants-conditions.xlsx* file.

Overview of Condition Sequencing Across All Participants

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1st	1	6	7	7	2	4	5	1	5	2	3	1	2	1	1	7	3	7	6	5	6	1
2nd	6	5	2	6	5	7	2	6	1	6	7	3	1	2	6	2	6	2	5	1	5	5
3rd	5	7	1	3	3	3	4	3	4	5	4	7	6	7	2	6	5	4	1	7	3	4
4th	7	3	5	5	6	1	3	5	6	7	2	2	5	3	3	3	1	5	7	3	4	7
5th	3	4	6	1	7	2	7	4	2	1	5	5	3	5	5	1	2	3	3	4	1	6
6th	4	2	3	2	4	6	6	2	3	4	1	6	7	4	7	5	4	6	2	6	2	3
7th	2	1	4	4	1	5	1	7	7	3	6	4	4	6	4	4	7	1	4	2	7	2

Table A.3: Presentation Order of Experimental Conditions for Each Participant. The horizontal axis displays participant numbers, while the vertical axis indicates the order in which conditions were presented to each participant.

A.5. Performance

A.5.1. Performance Across All Conditions

In the described methodology, performance is assessed by calculating the average items collected per minute throughout the entire condition. However, this metric does not provide insights into the participant's error rate. There exists a potential trade-off between speed and accuracy [3], suggesting that participants might make errors while aiming for quicker product collection.

Figure A.4 depicts a graph illustrating both speed (items per minute) and accuracy (number of errors in product selection) across all conditions. Instances, where participants skipped an item on the list or collected the wrong product, are considered errors, with the count including both types. In some cases, participants initially skipped an item but later realized the omission and collected the previous one, constituting a single error.

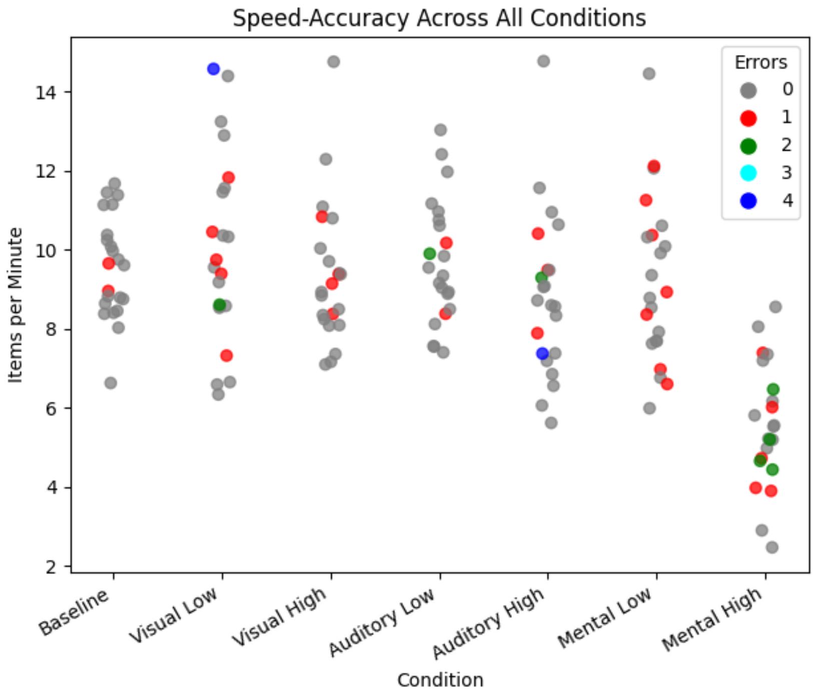


Figure A.4: Performance across all conditions with the condition on the horizontal axis, the items collected per minute (speed) on the vertical axis, and the number of errors made in picking products (accuracy) displayed by the color of the marker.

It could be observed that in every condition participants make errors. The majority of participants encountered a single error, except in the mental high condition, where the majority made two errors. In both the visual low and auditory high conditions, two distinct participants each made four errors. Notably, the participant in the visual low condition skipped four items, resulting in the recorded errors. It is essential to acknowledge that this participant managed to complete the list within the allocated time by skipping four items. However, the participant finished the session upon releasing the product, and therefore it is assumed that it did not have an impact on the results. In the auditory high condition, another participant skipped two items on the list twice.

A.5.2. Performance During The N-back Task

In the mental stimuli conditions, a secondary task—the n-back task—was introduced alongside the primary task. It's important to know that participants were explicitly instructed that both tasks had the same priority. Consequently, it remains uncertain how participants would prioritize the tasks in case of failure. Forster et al. [4] propose that individuals' speed and accuracy in task performance are influenced by their regulatory focus, whether emphasizing promotion (achieving gains, i.e., speed in this context) or prevention (avoiding losses, i.e., making no errors in this context). Consequently, depending on how participants interpret the task, performance may exhibit varied patterns across subjects.

It's noteworthy that none of the participants made errors in correctly repeating digits during the zero-back task in the mental low condition. Thus, the performance metric solely measures the speed of product picking. However, in the 2-back task, no participants managed to avoid errors, resulting in the creation of a graph exclusively for this task.

Figure A.5 depicts the score of the 2-back task on the horizontal axis and the speed in items per minute on the vertical axis. An interesting observation is that some participants perform well in both tasks, while others perform poorly in both. Additionally, it can be perceived that certain participants prioritize speed over accuracy, while others exhibit the opposite tendency. This variation may arise from the participants' interpretations, given the freedom in prioritization by themselves.

Examining the relationship between the performance in the main task and the 2-back task, as illustrated in Figure A.5, reveals a positive Pearson's correlation coefficient ($r = 0.483$) [5]. The observed correlation highlights the individualized nature of participants' responses to the mental high condition and emphasizes the influence of within-subject differences on task performance. While these results suggest potential nuances in how participants navigate the mental high condition, further investigation and analysis are required to draw definitive conclusions about the impact of individual variations on overall task performance.

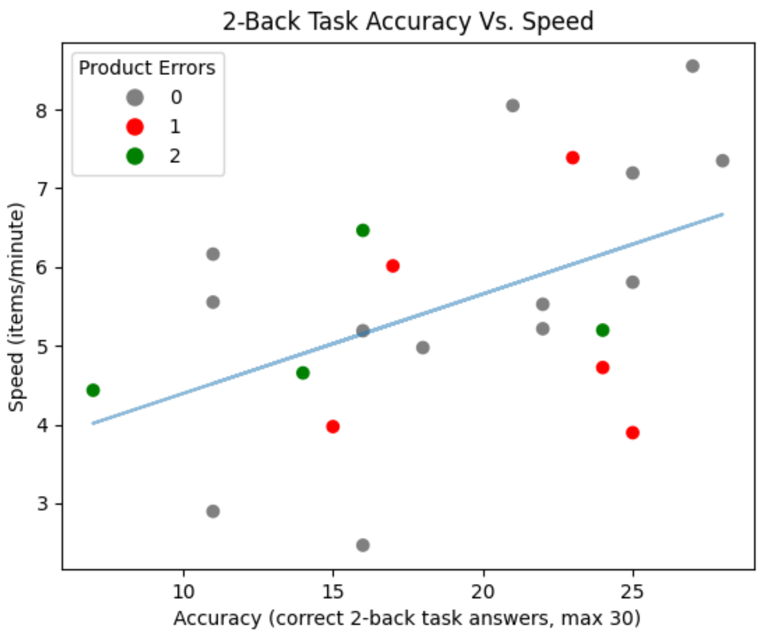


Figure A.5: The score of the 2-back task (max 30, on the horizontal axis) plotted against the performance of the main task (items per minute on the vertical axis). The blue line represents the correlation between the data points ($r = 0.483$). The colors in the figure represent the errors that subjects made in the main task (e.g. skipping an item or picking the wrong item).

A.6. Heart Rate

A.6.1. Outlier Detection

In order to analyze the heart rate, peak detection was performed. With the indices of the detected peaks, including the corresponding time intervals the heart rate and the variability could be determined. Before evaluating the results and computing the heart rate and variability for all participants, outliers were removed. This was achieved using interquartile range (IQR) outlier detection [6]. The IQR is calculated as the range between the 25th and 75th percentile sample quantiles, denoted as Q1 (lower quartile value) and Q3 (upper quartile value). Therefore the IQR equals $Q_3 - Q_1$. Data points outside the lower whisker limit ($Q_1 - 1.5IQR$) and higher whisker limit ($Q_3 + 1.5IQR$) are considered as outliers. For each condition and every subject, this outlier detection method was performed on the analyzed ECG signal. For a better understanding, pseudo-code is added in algorithm 1.

Algorithm 1 Delete Outliers using IQR

function DeleteOutliersIQR(detected_rr_peaks)
  $rr\_intervals \leftarrow diff(detected\_rr\_peaks)$
  $heart\_rate \leftarrow 60/(rr\_intervals/sample\_rate)$
  $sorted\_data \leftarrow sort(heart\_rate)$
  $quartile\_1 \leftarrow sorted\_data[\lfloor \text{len}(heart\_rate) \times 0.25 \rfloor]$
  $quartile\_3 \leftarrow sorted\_data[\lfloor \text{len}(heart\_rate) \times 0.75 \rfloor]$
  $iqr \leftarrow quartile\_3 - quartile\_1$
  $lower\_bound \leftarrow quartile\_1 - 1.5 \times iqr$
  $upper\_bound \leftarrow quartile\_3 + 1.5 \times iqr$
  $filtered\_heart\_rate \leftarrow$ [if $lower\_bound \leq x \leq upper\_bound$ then x else np.nan for x in $heart\_rate$]
  return $filtered\_heart\_rate$
end function

A.6.2. Signals Over Time

The paper has previously discussed the fluctuation of heart rate in both the mental low and high conditions over time. The average heart rate across all participants for all seven conditions can be found in Figure A.6 (similar to figure in the paper). This was achieved by first establishing a common x-axis through linear interpolation between the minimum and maximum time values. Following this, the filtered peak values corresponding to individual time series were interpolated to align with this common x-axis. The mean of the interpolated values along the vertical axis, representing heart rate, was then calculated for each corresponding x-value. This process generates a smooth average signal that captures the central tendency of the heart rate data over time for the specified experimental conditions.

What can be observed in the figure is that the average heart rate for all conditions starts at approximately the same value of beats per minute (ranging between 78.1 and 83.1 beats/minute). Similarly, the mean heart rate at the end of the conditions is approximately identical (ranging between 85.0 and 88.9). However, what stands out is that, in the first minute, the average heart rates for some conditions show an increase, particularly clear in the mental low and mental high conditions. In the mental low condition, the heart rate increases from 82.8 to 96.9 beats/minute within 17.1 seconds. Simultaneously, the average heart rate for the mental high load condition rises from 83.1 to 103.3 beats/minute within 22.0 seconds. Following the rapid increase, the average heart rates for these conditions subsequently decrease. Additionally, for the conditions without stimuli, namely the visual low and auditory high conditions, the increase is most pronounced at the start at the same rate. In the visual high and auditory low conditions, the average heart rate also increases, but at a slower pace compared to the other conditions.

According to literature, an increase in mean heart rate is expected to correspond to a decrease in mean heart rate variability [7]. Therefore, a similar technique has been employed to visualize heart rate variability. Similar to the heart rate, it could be observed that heart rate variability decreases for both low and high mental load conditions in comparison with the other stimuli. In Figure A.7, the

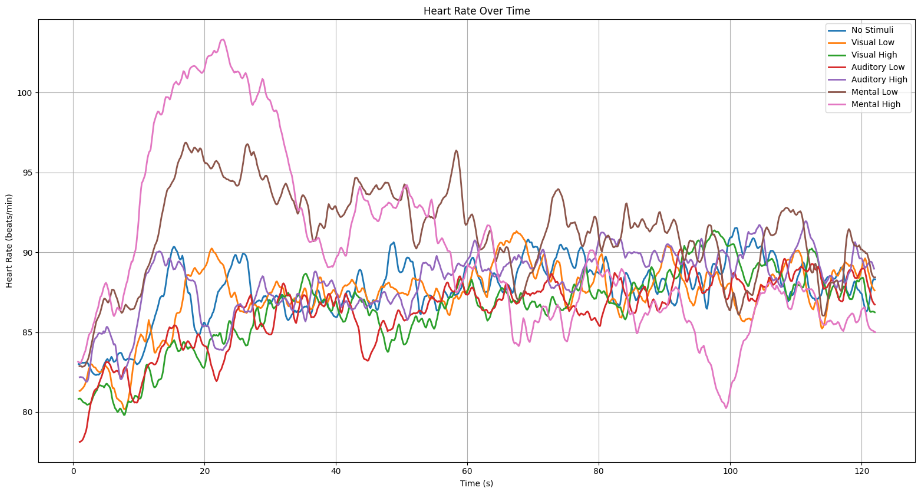


Figure A.6: The average heart rate over time for all conditions.

mental stimuli are outlined, highlighting a clear trend toward lower heart rate variability within the first 40 seconds.

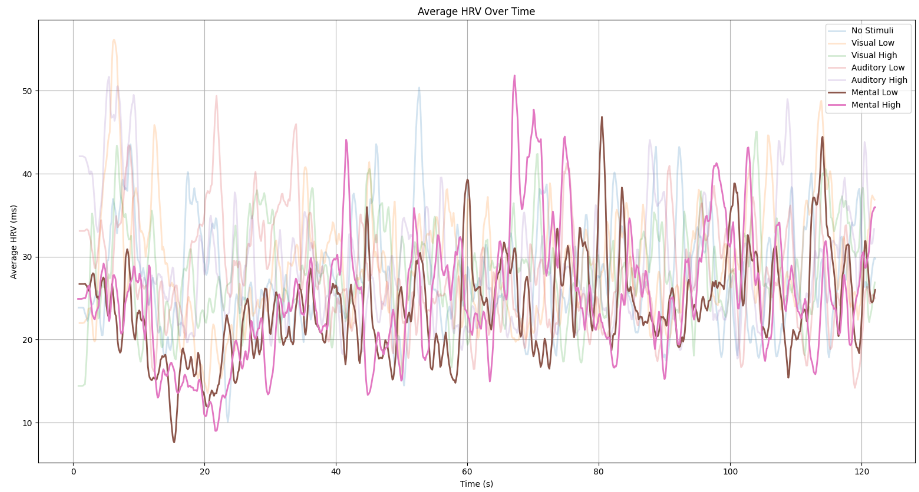


Figure A.7: The average heart rate variability over time for all conditions, but the mental conditions are outlined.

It is important to note that these visualizations offer a general overview of heart rate and its variability over time. While they provide some insights into the signal's course, further investigation is necessary to derive proper conclusions from it. These graphs rely solely on averages and lack insights into the standard deviation. Moreover, the use of IQR outlier detection may have led to the exclusion of true positive peaks, impacting the analyzed graphs and potentially yielding different outcomes. The graphs also overlook the order of conditions, as the experiment did not measure the heart rate recovery to baseline. This absence of recovery time increases the possibility that other conditions may have influenced the heart rate in subsequent conditions. Lastly, it is important to acknowledge the limitation in the sample size, as only data from 22 subjects were considered.

A.7. Area of Interest

As illustrated in the scientific paper, there are differences in some conditions compared to the baseline regarding the gaze location. Although the results do not directly indicate that there exists an increase in cognitive load, it might still be relevant for future research. For example, even though the distraction may not influence cognitive load in healthy subjects, it may have a larger influence on patients in neurorehabilitation with cognitive deficits. Therefore, this section delves more into the area of interest, and particularly in the effects of the features of the visual stimuli.

Figure A.8 is a copy of what has been shown in section A.1, namely the area of interest on other objects than the main shelf, the *shopping list*, or the *cart*. What catches the eye is that in the visual conditions, the total gaze duration on *other objects* is higher. Indeed, this makes sense since these are the conditions where there are visual distractions present in the environment. After further inspection, it turns out that from the total time looking at *other objects*, 66.81%, and 61.76% of the time in the visual low and visual high conditions respectively are spent allocating gaze location at a character (NPC). In the other conditions, these calculations could not be performed since there were simply no NPCs present.

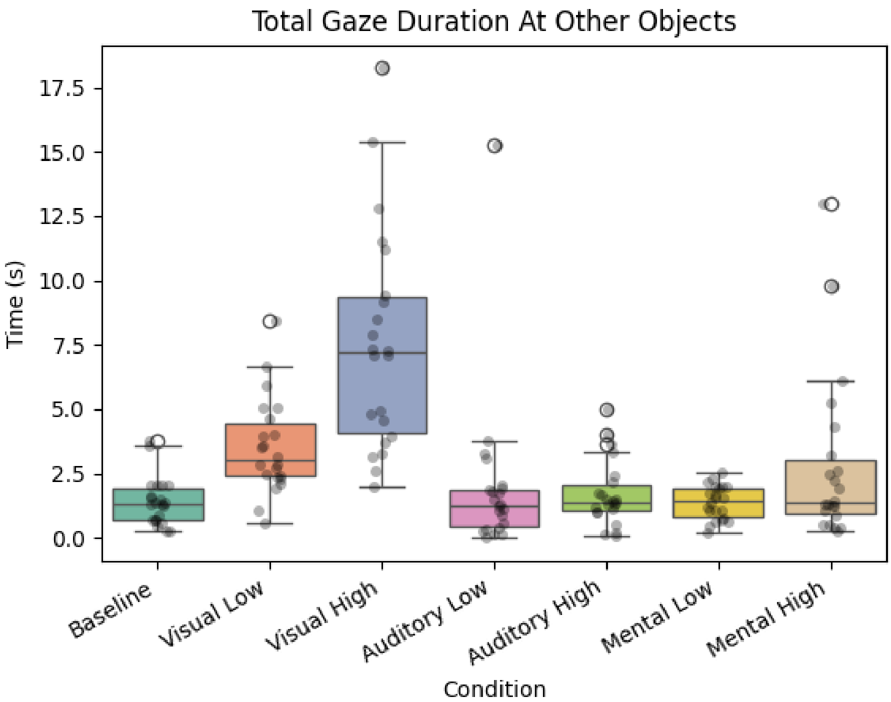


Figure A.8: Box plot showing total time gazing at *other objects* across conditions.

It should be mentioned that the gaze duration on *other objects* covers a small fraction of the time compared to the *shopping list, cart*, and *main shelf*. Subjects were mainly focused on executing the main task and little time was spent to focus on something else. For an overview of the percentile distribution of gaze location across the conditions can be found in figure A.9.
To determine the location where the participants were looking, the gaze locations have been calculated. Three components were extracted from the eyes using the HMD: the origin of the ray (rayOrigin), the direction of the ray (rayDirection), and the convergence distance of the eyes (convergenceDistance). The convergence distance is the distance from the head-mounted display where the rays of the two eyes cross. This is not always accurate, so the gaze location is not always precise. However, the ray origin and direction are very accurate according to the software. Therefore, if a heat map is created, it should give an approximate idea of where subjects were looking in the environment. This section only shows where subjects were looking when they were focusing on other objects instead of the main task, to determine which aspects of the visual stimuli gain attention. The gaze locations for *other objects* are

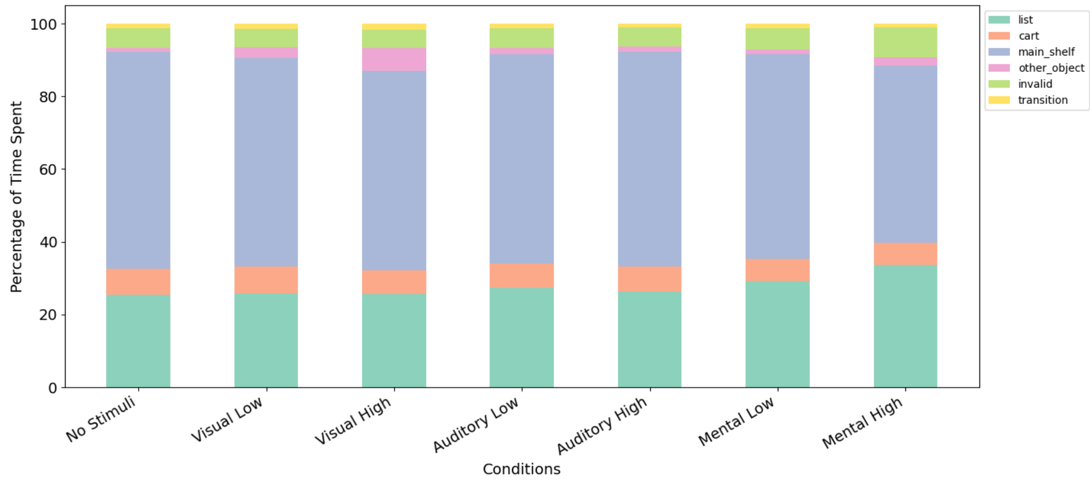
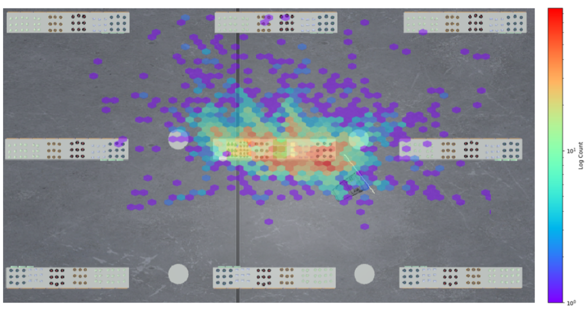


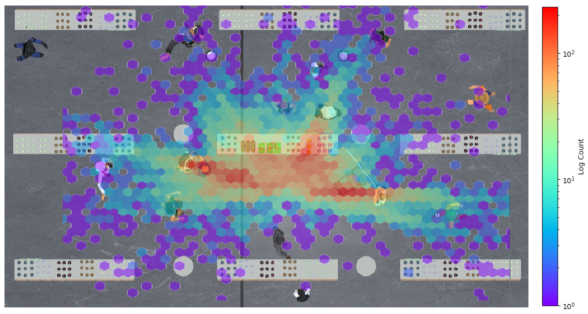
Figure A.9: Percentile distribution of gaze location across the conditions. The horizontal axis shows the name of the condition, and the vertical axis shows the percentage of the condition spent on a certain AOI.

plotted in a heat map in the XZ plane (top view) in Figure A.10. The gaze location is calculated using the following formula: $gazeLocation = rayOrigin + rayDirection \cdot convergenceDistance$.

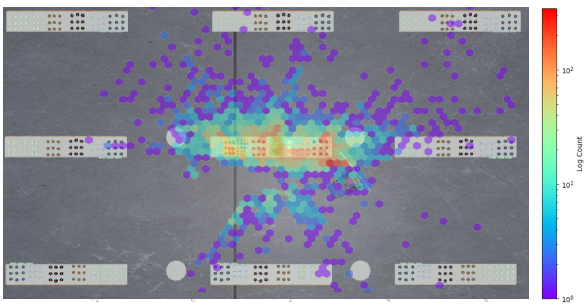
The top view heat maps are projected on the Unity map to give an idea of where participants focused the most. To show the difference between the conditions where almost no time was spent at *other objects*, the heat maps of the baseline, visual high, auditory high, and mental high are shown in figure A.10. Note that the color scales are different in each condition.



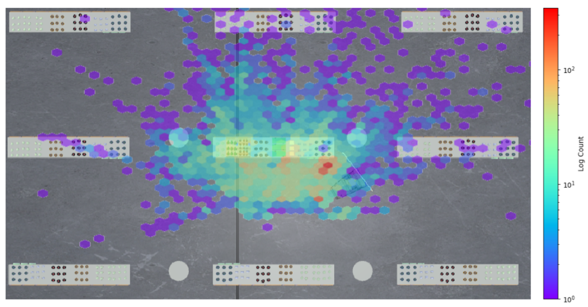
(a) Heat map of the gaze locations when focusing on *other objects* than the main task, baseline condition



(b) Heat map of the gaze locations when focusing on *other objects* than the main task, visual high condition



(c) Heat map of the gaze locations when focusing on *other objects* than the main task, auditory high condition



(d) Heat map of the gaze locations when focusing on *other objects* than the main task, mental high condition

Figure A.10: Heat maps of the gaze locations when focusing on *other objects* than the main task. The color scales are different in each condition.

What could be observed is that especially in the visual high condition, some ray directions are frequently visited. Mainly the woman and child left to the main shelf (**1** in figure A.11), the man behind the main shelf (**2** in figure A.11) and the woman right to the list (**3** in Figure A.11) are frequently visited by the subjects. Furthermore, if we then compute the distribution of time spent on the NPCs, it turns out that these three NPC locations cover 68.23% of the total gaze focus on NPCs.

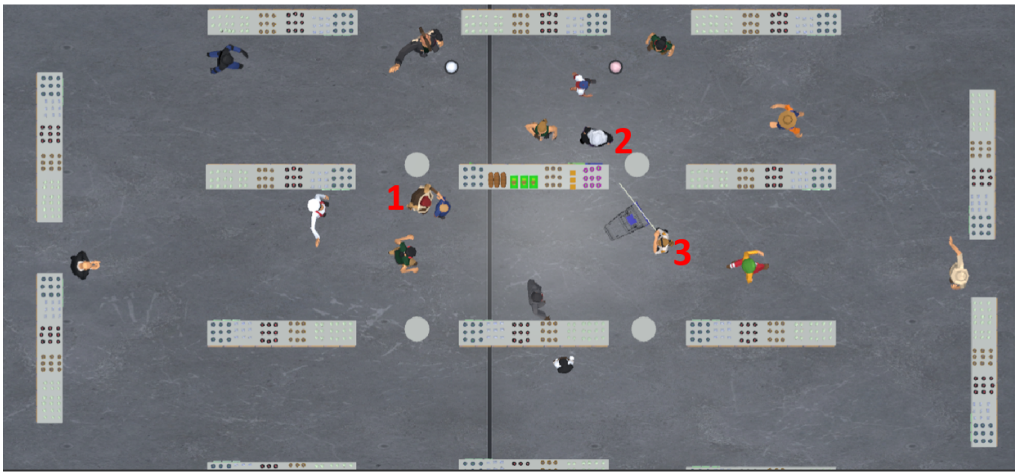


Figure A.11: Top view of the visual high condition including the idle and some of the moving characters. The three numbers (1 to 3) are located at the characters that are the most visited by the subjects.

The arrangement of the visual stimuli was conceived based on four elements: the type of character, the actions performed by the character, external events, and the positioning of the visuals (for details see Appendix B). Considering the results regarding the gaze locations when looking at *other objects*, I would say that the location of the visual stimuli exerts the greatest influence on distracting the participants visually. Characters that are visible and close to the participant while they perform the main task, engaging in interactions such as waving, moving, or smiling toward the participant, appear to attract the most attention. Notably, the three most frequently visited characters exhibited distinct actions, and the introduced external event (flashing light) did not capture the attention of any participants.

A.8. Behavior Charts

The presented scientific paper illustrates 1 of the 22 behavior charts (participant 4), but variations among participants exist. This section displays the behavior charts containing all conditions across all participants, in Figures A.12 to A.33.

A.9. Additional Statistical Analysis Results

As mentioned in the scientific paper, more statistical effect sizes were calculated. Due to space constraints, only an alternative calculation for Cohen's d was added to the paper. The calculations of the alternative Cohen's d and d_z were performed by sorting based on the condition order per participant, followed by z-score transformation, and then reverted to its original order. For convenience, this section contains the values for the conventional Cohen's d and d_z and the values for the alternative Cohen's d and d_z in Table A.4. The reported value for Cohen's d and d_z is highlighted in bold text if the corrected p-value was significant ($p < 0.05$).

Additional Results from the Statistical Analysis

Measure	**Visual Low**	**Visual High**	**Auditory Low**	**Auditory High**	**Mental Low**	**Mental High**
Cohen's d, without correction for the condition order						
Gaze Duration on *Cart*	0.04	-0.26	-0.28	-0.13	-0.59	-0.52
Gaze Duration on *Shopping List*	0.10	0.05	0.46	0.18	**0.74**	**1.33**
Gaze Duration on *Main Shelf*	-0.45	-0.97	-0.36	-0.12	-0.57	**-1.52**
Gaze Duration on *Other Objects*	**1.44**	**1.98**	-0.04	0.26	-0.04	0.58
Mean Heart Rate	0.00	-0.08	-0.07	0.04	0.28	0.20
Mean Heart Rate Variability	0.05	0.04	-0.02	0.07	-0.28	-0.05
Hand Movement Smoothness	0.21	0.13	0.21	0.12	0.18	0.22
Head Stillness Duration	0.05	-0.41	0.03	0.06	0.11	**1.14**
Performance	0.23	-0.13	0.09	-0.43	-0.21	**-2.80**
Weighted NASA-TLX Score	-0.07	0.44	0.19	**1.03**	**1.33**	**3.71**
Cohen's d, with correction for the condition order (values similar to the paper)						
Gaze Duration on *Cart*	0.01	-0.33	-0.31	-0.18	-0.70	-0.52
Gaze Duration on *Shopping List*	0.05	-0.08	0.24	0.15	**0.74**	**1.31**
Gaze Duration on *Main Shelf*	-0.48	-0.72	-0.18	0.01	-0.59	**-1.48**
Gaze Duration on *Other Objects*	**1.32**	**2.08**	0.14	0.15	-0.12	0.53
Mean Heart Rate	0.01	-0.06	-0.05	0.05	0.27	0.20
Mean Heart Rate Variability	-0.07	-0.01	0.00	0.00	-0.30	-0.06
Hand Movement Smoothness	0.07	0.06	-0.04	0.10	0.11	0.13
Head Stillness Duration	0.13	-0.31	0.32	0.01	0.20	**1.11**
Performance	0.09	-0.26	-0.39	-0.38	-0.25	**-3.21**
Weighted NASA-TLX Score	0.03	0.55	0.41	**1.07**	**1.45**	**4.82**
Cohen's d_z, without correction for the condition order						
Gaze Duration on *Cart*	0.04	-0.31	-0.36	-0.16	-0.64	-0.42
Gaze Duration on *Shopping List*	0.17	0.06	0.52	0.27	**0.86**	**0.97**
Gaze Duration on *Main Shelf*	-0.47	-0.94	-0.36	-0.13	-0.68	**-1.14**

Table A.4 – Continued from previous page

Measure	**Visual Low**	**Visual High**	**Auditory Low**	**Auditory High**	**Mental Low**	**Mental High**
Gaze Duration on *Other Objects*	**1.21**	**1.52**	-0.04	0.27	-0.04	0.42
Mean Heart Rate	0.00	-0.21	-0.20	0.10	0.75	0.54
Mean Heart Rate Variability	0.11	0.08	-0.04	0.14	-0.50	-0.08
Hand Movement Smoothness	0.27	0.18	0.27	0.19	0.25	0.23
Head Stillness Duration	0.06	-0.49	0.04	0.08	0.13	**1.00**
Performance	0.26	-0.13	0.09	-0.46	-0.19	**-2.15**
Weighted NASA-TLX Score	-0.10	0.51	0.29	**1.04**	**1.19**	**2.47**
Cohen's d_z, with correction for the condition order						
Gaze Duration on *Cart*	0.01	-0.42	-0.43	-0.24	-0.81	-0.44
Gaze Duration on *Shopping List*	0.11	-0.09	0.28	0.22	**0.93**	**0.97**
Gaze Duration on *Main Shelf*	-0.52	-0.82	-0.18	0.01	-0.72	**-1.18**
Gaze Duration on *Other Objects*	**1.03**	**1.80**	0.13	0.14	-0.10	0.45
Mean Heart Rate	0.06	-0.16	-0.15	0.19	0.78	0.52
Mean Heart Rate Variability	-0.18	-0.02	-0.01	-0.01	-0.53	-0.11
Hand Movement Smoothness	0.13	0.09	-0.06	0.16	0.20	0.17
Head Stillness Duration	0.20	-0.54	0.44	0.03	0.26	**1.01**
Performance	0.14	-0.37	-0.56	-0.42	-0.26	**-2.33**
Weighted NASA-TLX Score	0.03	0.78	0.55	**1.02**	**1.53**	**3.66**

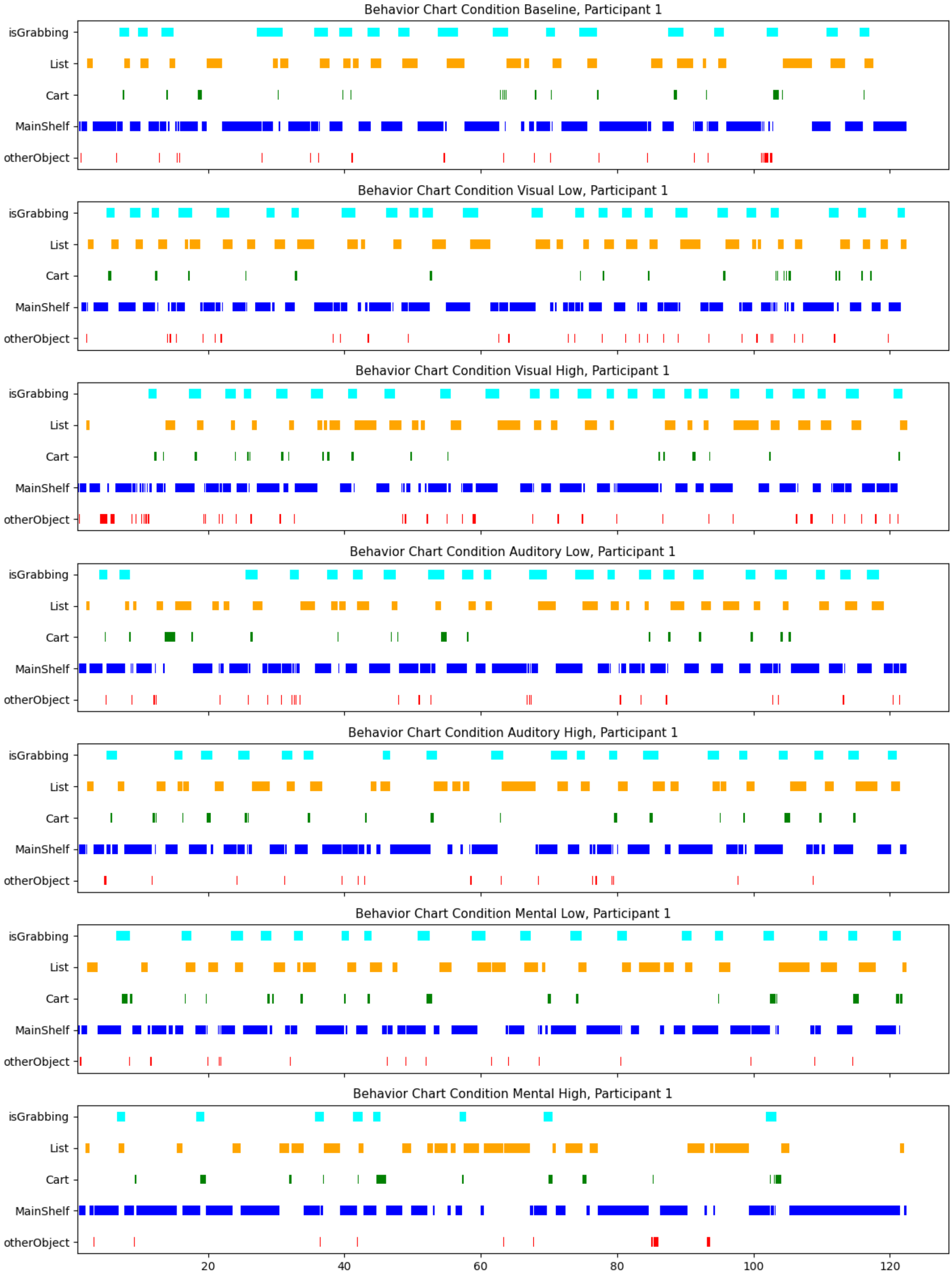


Figure A.12: The behavioral chart illustrates the gaze patterns and object interaction of Participant 1 across all conditions as detailed in the scientific paper. The vertical axis shows the areas of interest within the environment, while colored bars denote the participant's gaze allocation during each condition. Additionally, the presence of object grabbing is displayed. The horizontal axis represents the time, ranging from 0 to 122 seconds.

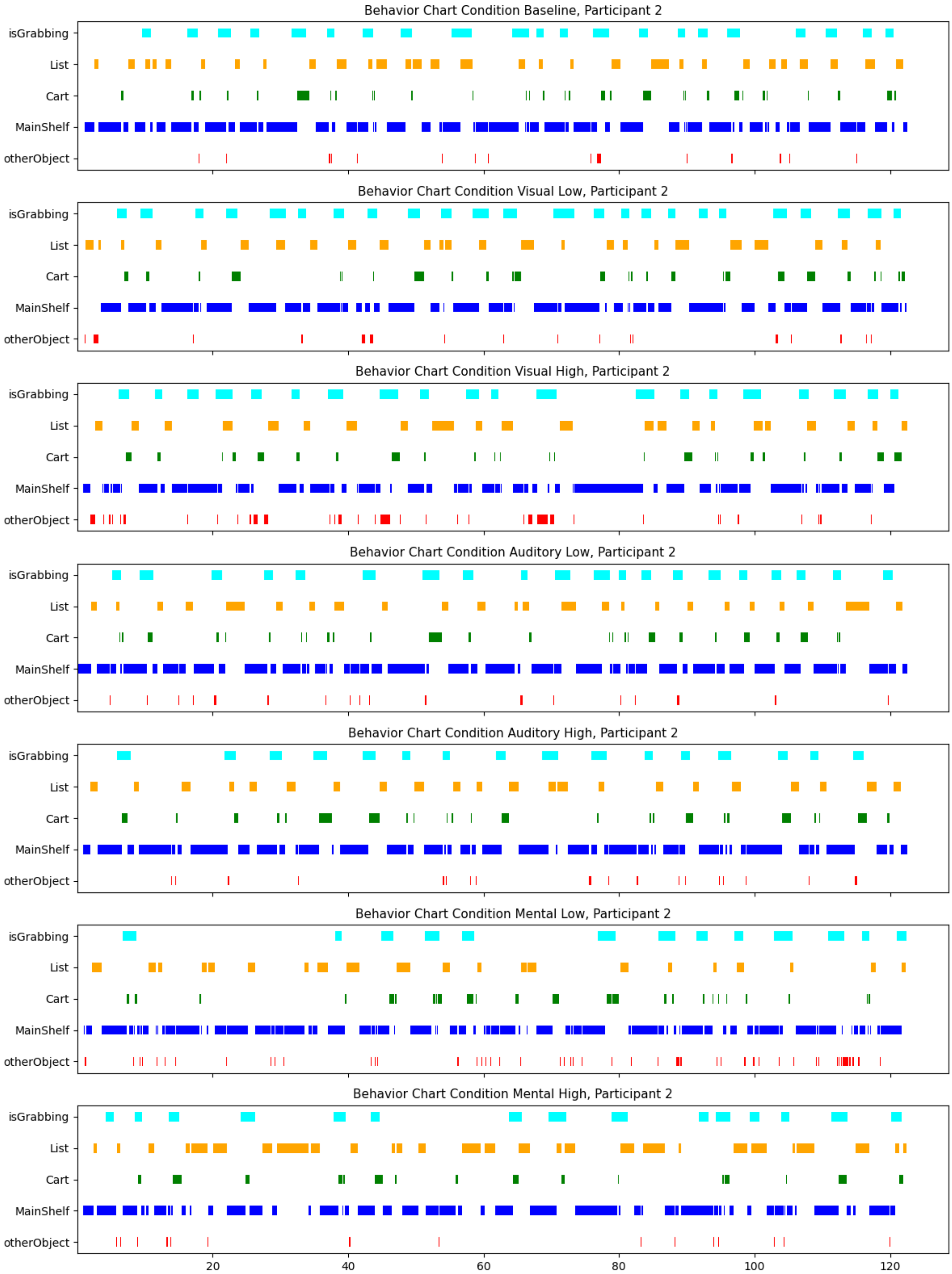


Figure A.13: The behavioral chart illustrates the gaze patterns and object interaction of Participant 2 across all conditions as detailed in the scientific paper. The vertical axis shows the areas of interest within the environment, while colored bars denote the participant's gaze allocation during each condition. Additionally, the presence of object grabbing is displayed. The horizontal axis represents the time, ranging from 0 to 122 seconds.

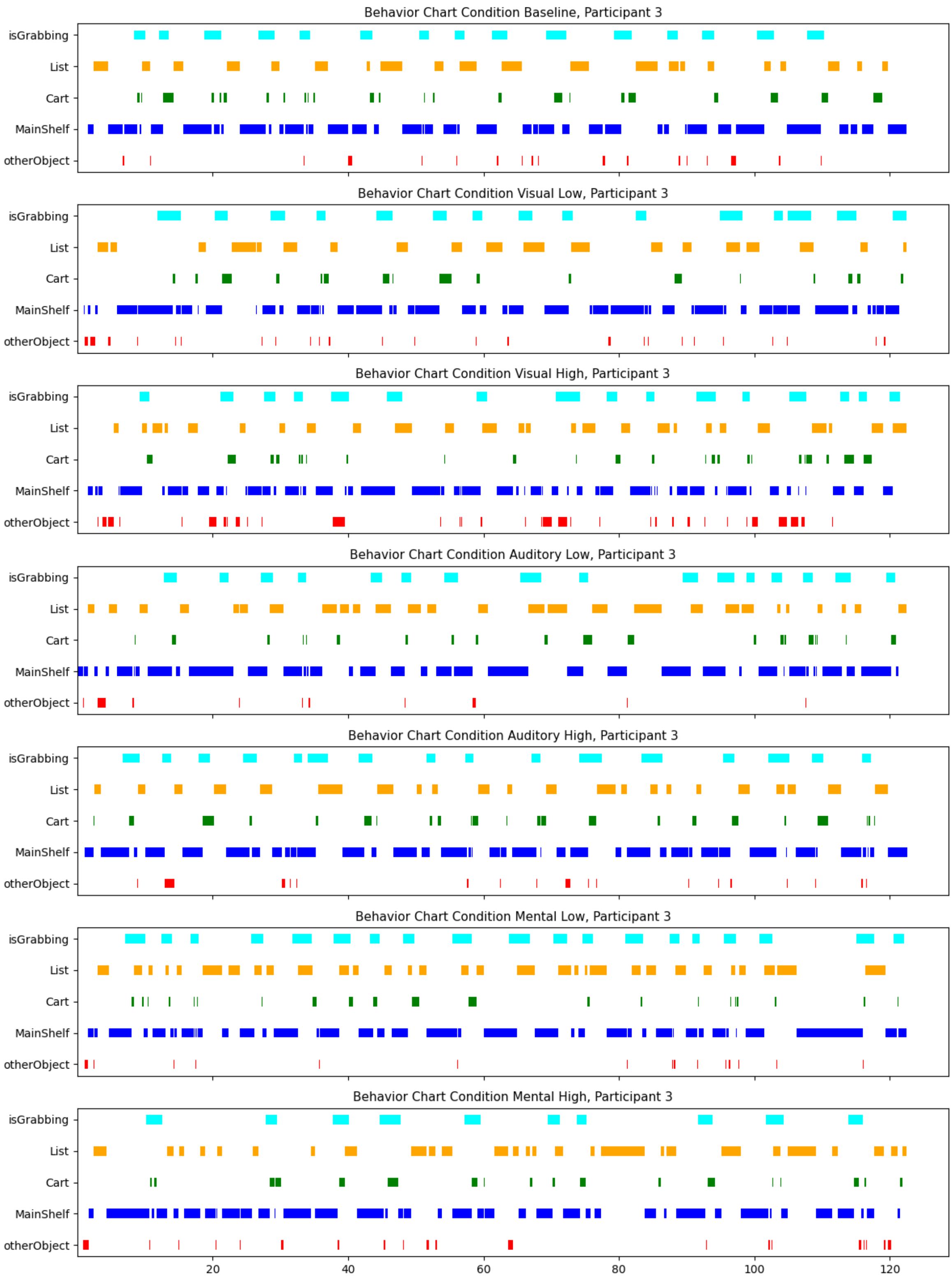


Figure A.14: The behavioral chart illustrates the gaze patterns and object interaction of Participant 3 across all conditions as detailed in the scientific paper. The vertical axis shows the areas of interest within the environment, while colored bars denote the participant's gaze allocation during each condition. Additionally, the presence of object grabbing is displayed. The horizontal axis represents the time, ranging from 0 to 122 seconds.

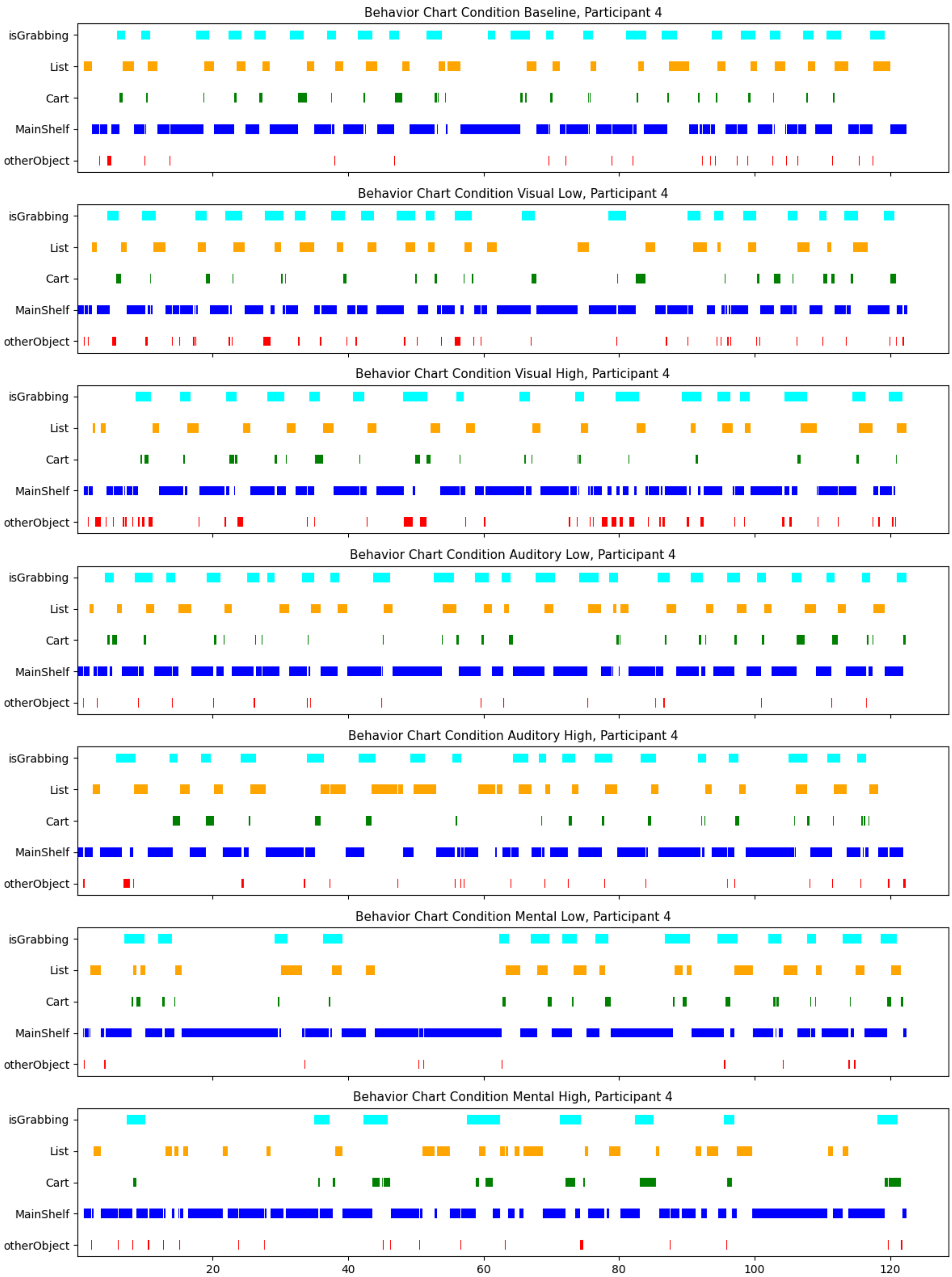


Figure A.15: The behavioral chart illustrates the gaze patterns and object interaction of Participant 4 across all conditions as detailed in the scientific paper. The vertical axis shows the areas of interest within the environment, while colored bars denote the participant's gaze allocation during each condition. Additionally, the presence of object grabbing is displayed. The horizontal axis represents the time, ranging from 0 to 122 seconds.

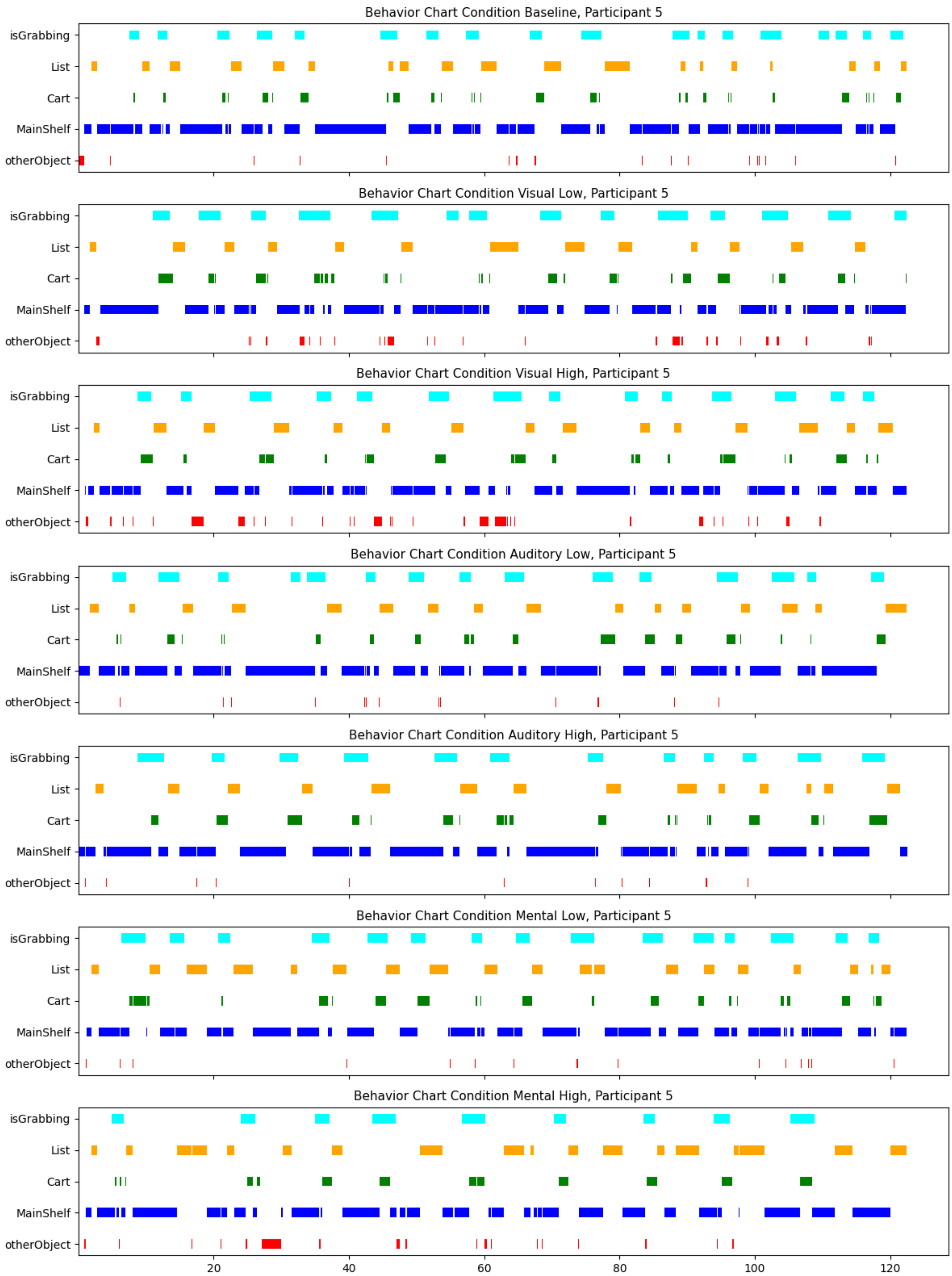


Figure A.16: The behavioral chart illustrates the gaze patterns and object interaction of Participant 5 across all conditions as detailed in the scientific paper. The vertical axis shows the areas of interest within the environment, while colored bars denote the participant's gaze allocation during each condition. Additionally, the presence of object grabbing is displayed. The horizontal axis represents the time, ranging from 0 to 122 seconds.

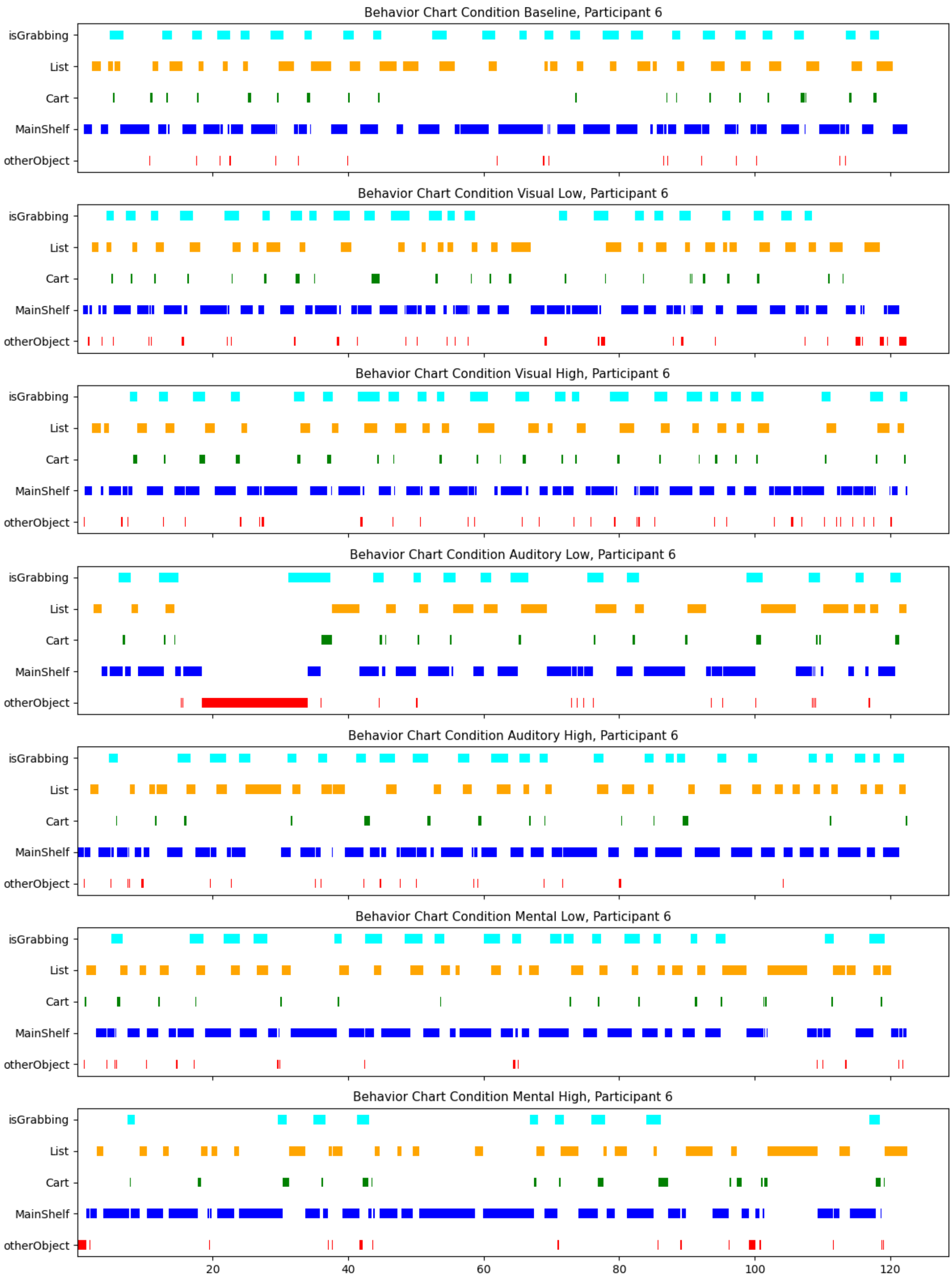


Figure A.17: The behavioral chart illustrates the gaze patterns and object interaction of Participant 6 across all conditions as detailed in the scientific paper. The vertical axis shows the areas of interest within the environment, while colored bars denote the participant's gaze allocation during each condition. Additionally, the presence of object grabbing is displayed. The horizontal axis represents the time, ranging from 0 to 122 seconds.

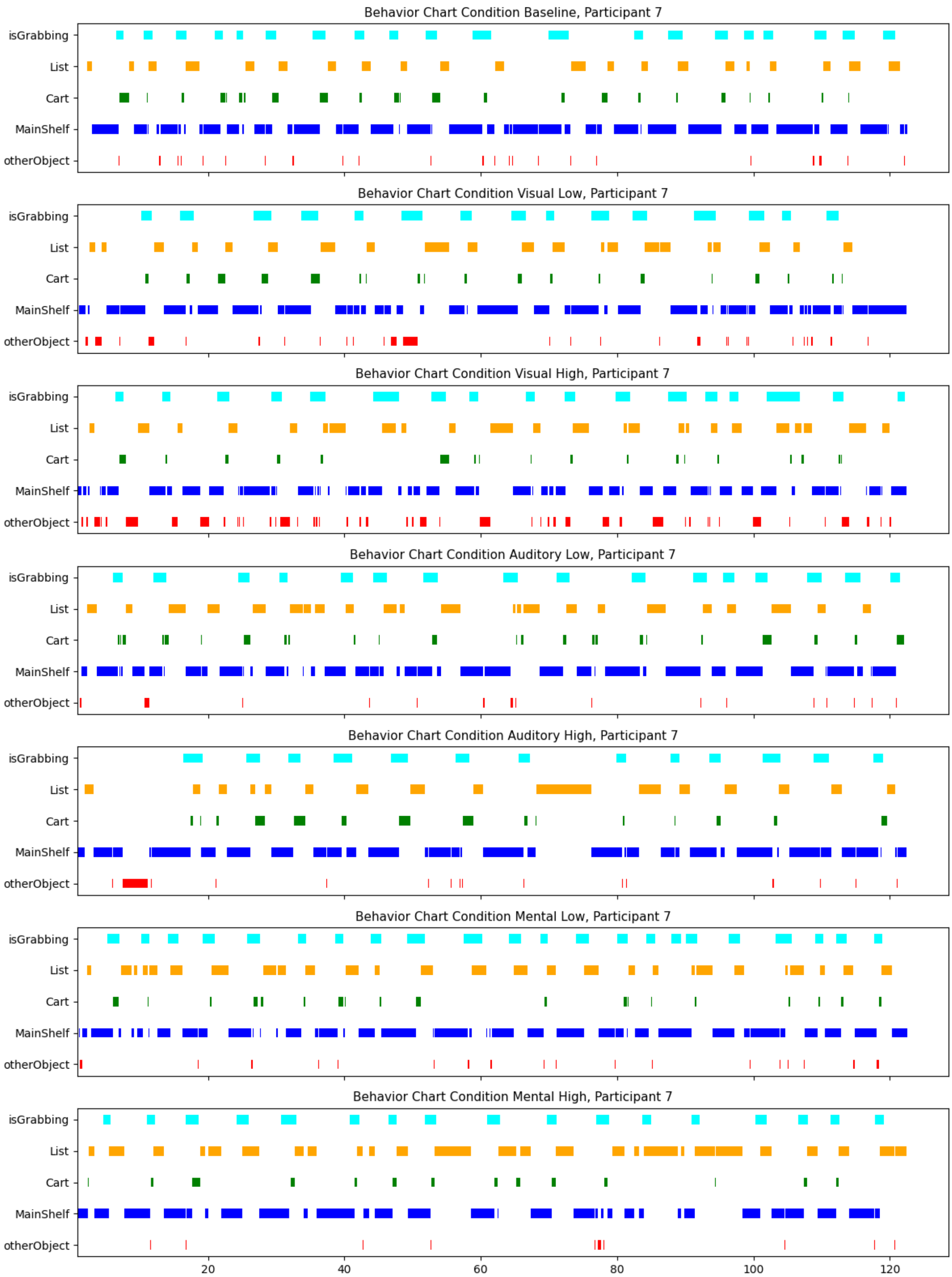


Figure A.18: The behavioral chart illustrates the gaze patterns and object interaction of Participant 7 across all conditions as detailed in the scientific paper. The vertical axis shows the areas of interest within the environment, while colored bars denote the participant's gaze allocation during each condition. Additionally, the presence of object grabbing is displayed. The horizontal axis represents the time, ranging from 0 to 122 seconds.

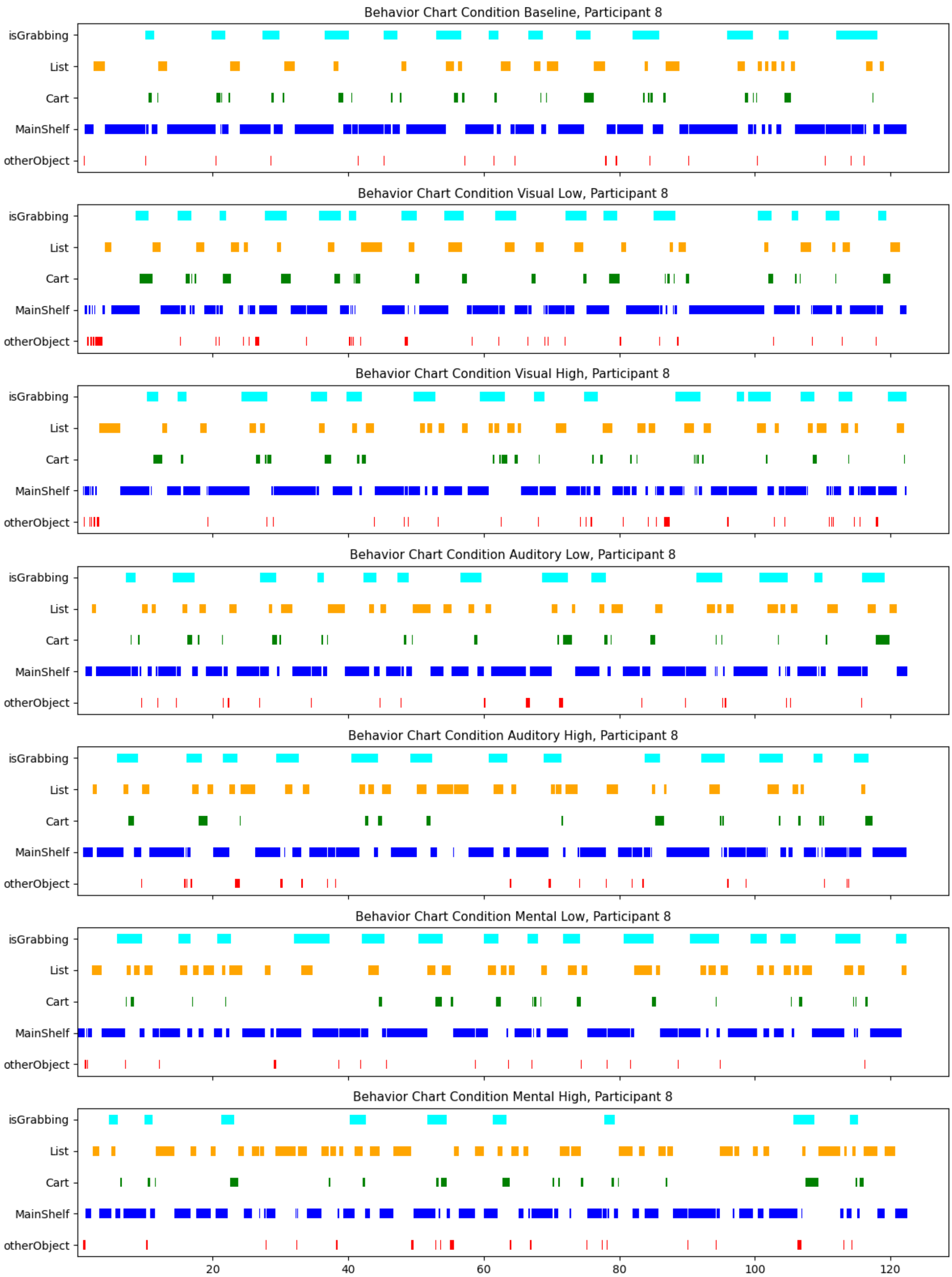


Figure A.19: The behavioral chart illustrates the gaze patterns and object interaction of Participant 8 across all conditions as detailed in the scientific paper. The vertical axis shows the areas of interest within the environment, while colored bars denote the participant's gaze allocation during each condition. Additionally, the presence of object grabbing is displayed. The horizontal axis represents the time, ranging from 0 to 122 seconds.

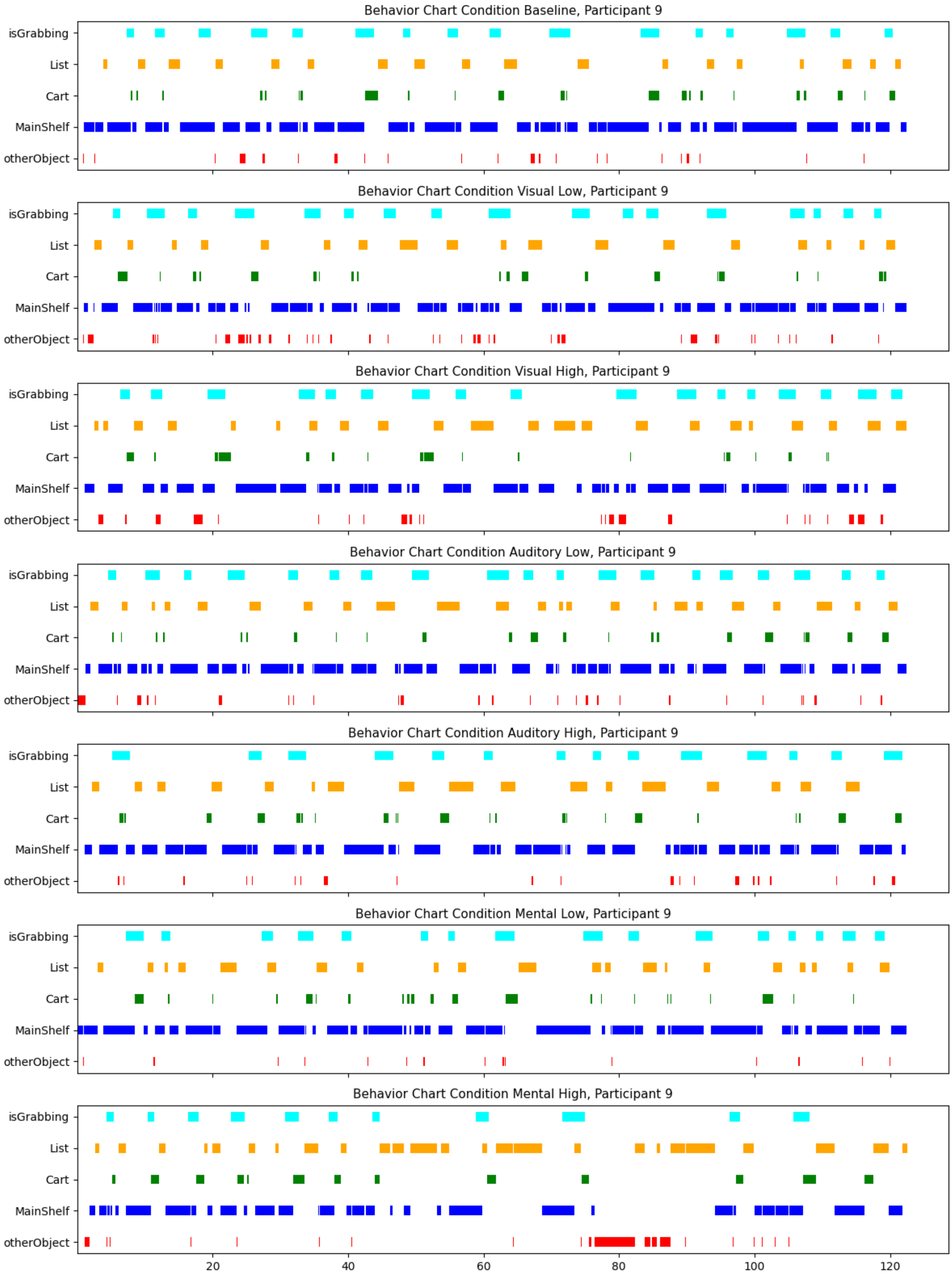


Figure A.20: The behavioral chart illustrates the gaze patterns and object interaction of Participant 9 across all conditions as detailed in the scientific paper. The vertical axis shows the areas of interest within the environment, while colored bars denote the participant's gaze allocation during each condition. Additionally, the presence of object grabbing is displayed. The horizontal axis represents the time, ranging from 0 to 122 seconds.

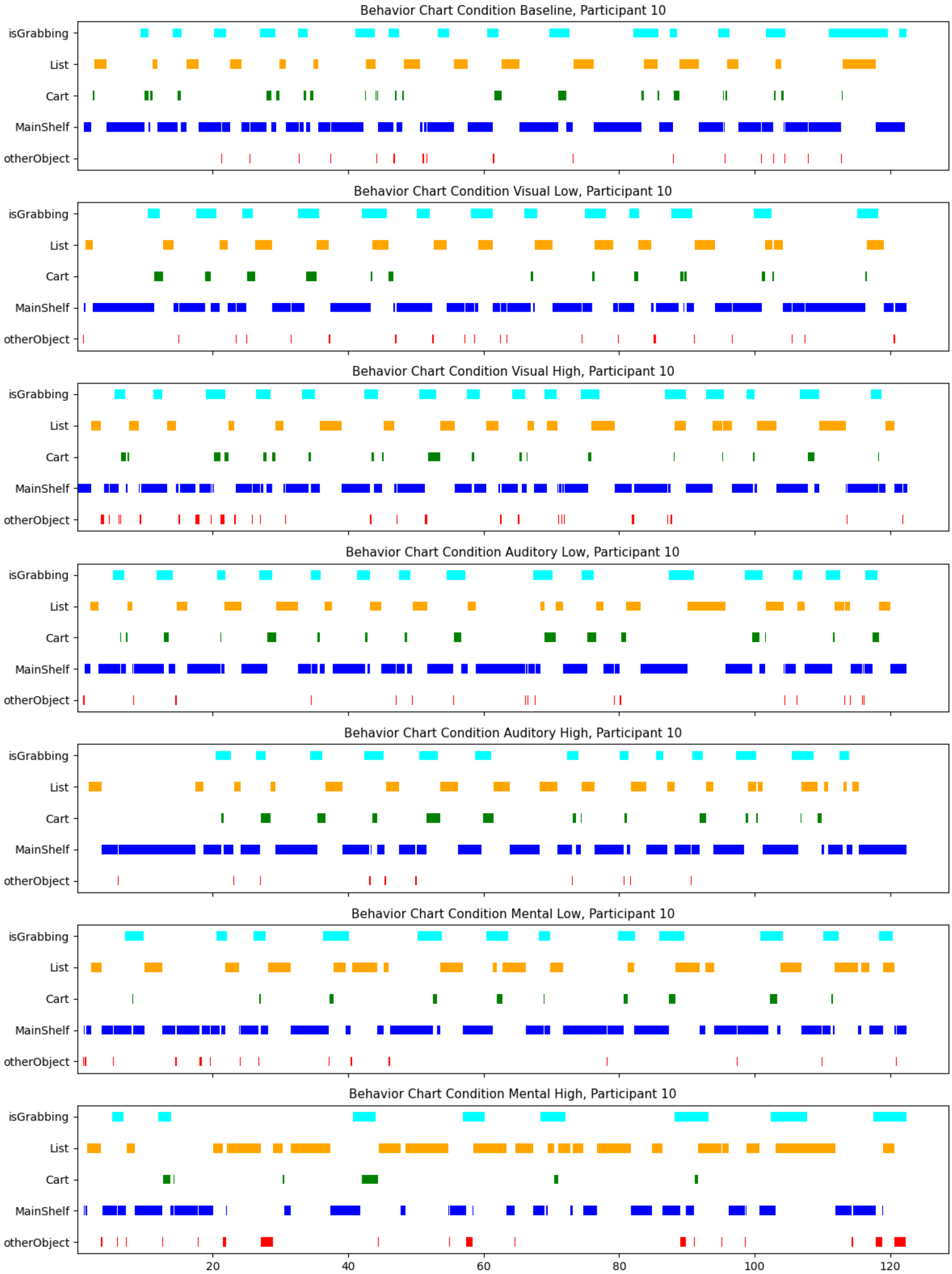


Figure A.21: The behavioral chart illustrates the gaze patterns and object interaction of Participant 10 across all conditions as detailed in the scientific paper. The vertical axis shows the areas of interest within the environment, while colored bars denote the participant's gaze allocation during each condition. Additionally, the presence of object grabbing is displayed. The horizontal axis represents the time, ranging from 0 to 122 seconds.

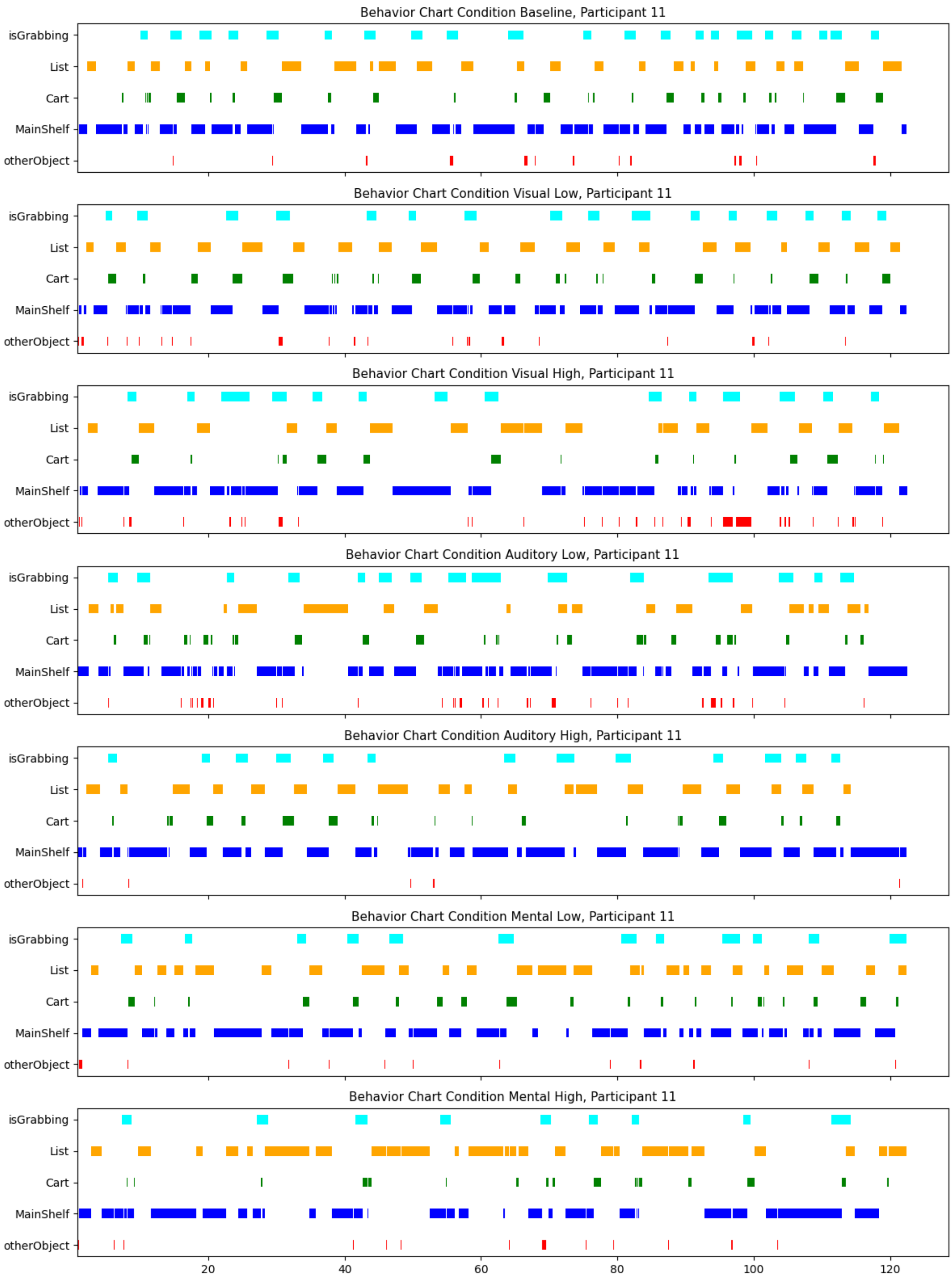


Figure A.22: The behavioral chart illustrates the gaze patterns and object interaction of Participant 11 across all conditions as detailed in the scientific paper. The vertical axis shows the areas of interest within the environment, while colored bars denote the participant's gaze allocation during each condition. Additionally, the presence of object grabbing is displayed. The horizontal axis represents the time, ranging from 0 to 122 seconds.

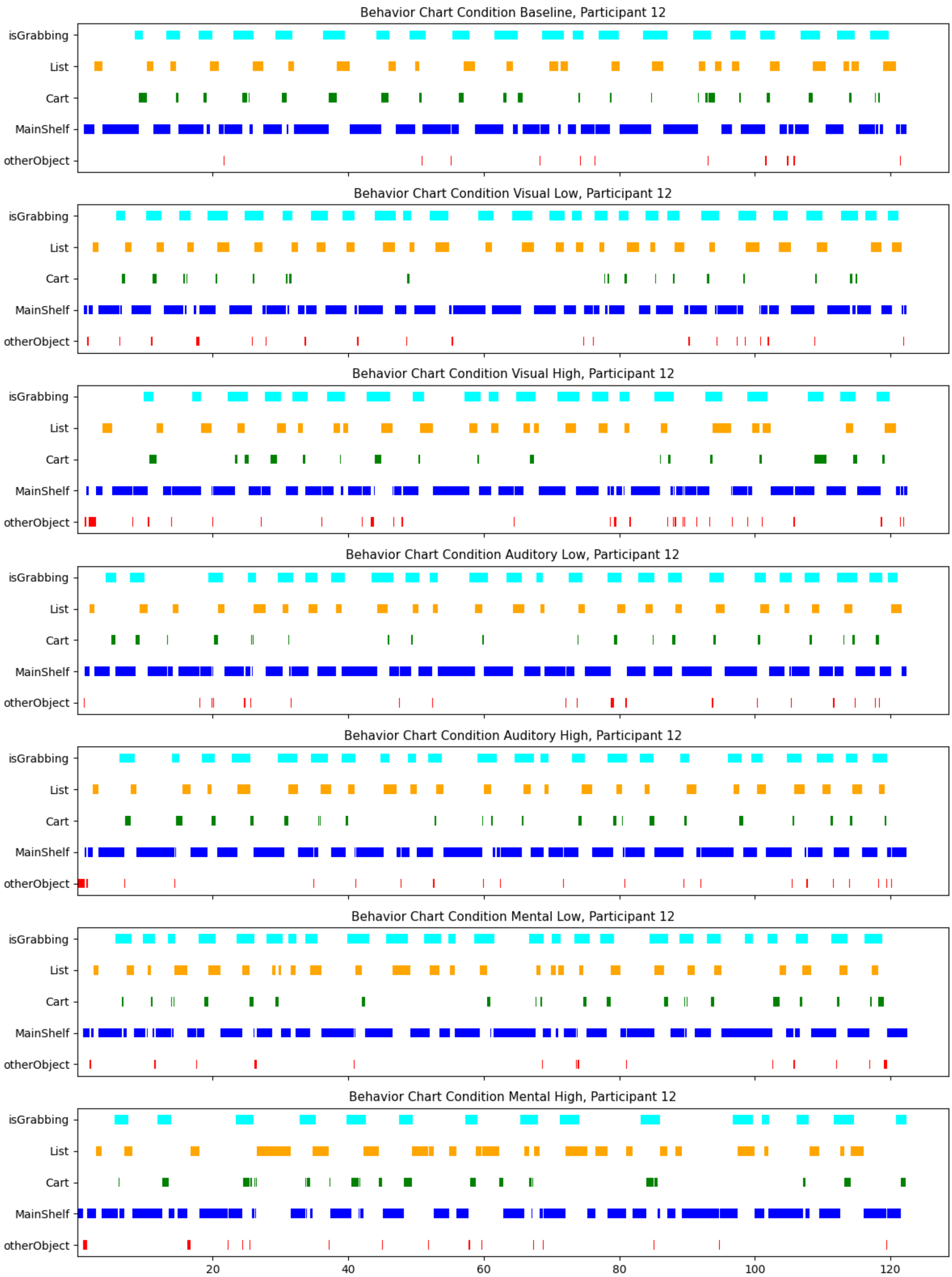


Figure A.23: The behavioral chart illustrates the gaze patterns and object interaction of Participant 12 across all conditions as detailed in the scientific paper. The vertical axis shows the areas of interest within the environment, while colored bars denote the participant's gaze allocation during each condition. Additionally, the presence of object grabbing is displayed. The horizontal axis represents the time, ranging from 0 to 122 seconds.

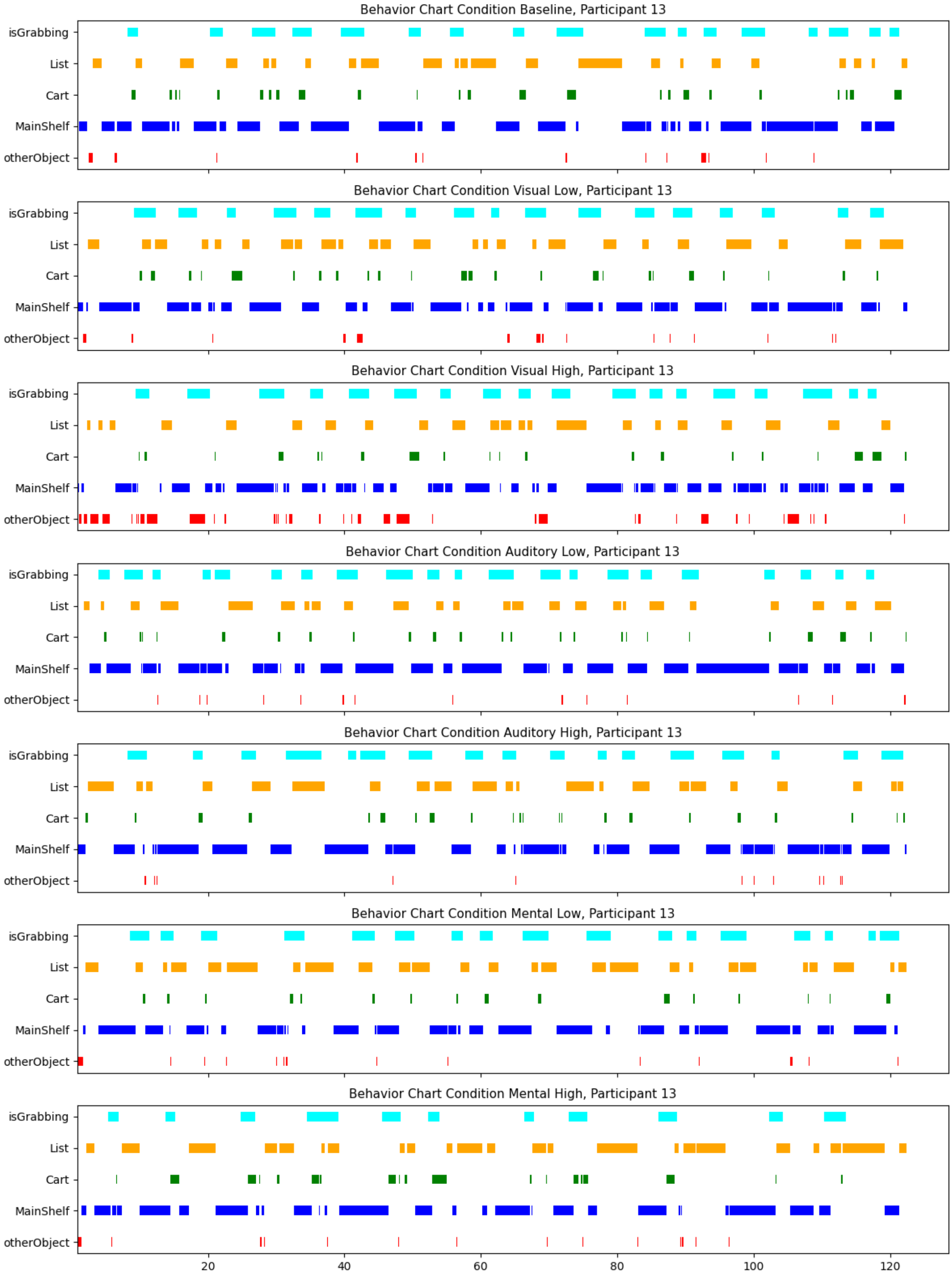


Figure A.24: The behavioral chart illustrates the gaze patterns and object interaction of Participant 13 across all conditions as detailed in the scientific paper. The vertical axis shows the areas of interest within the environment, while colored bars denote the participant's gaze allocation during each condition. Additionally, the presence of object grabbing is displayed. The horizontal axis represents the time, ranging from 0 to 122 seconds.

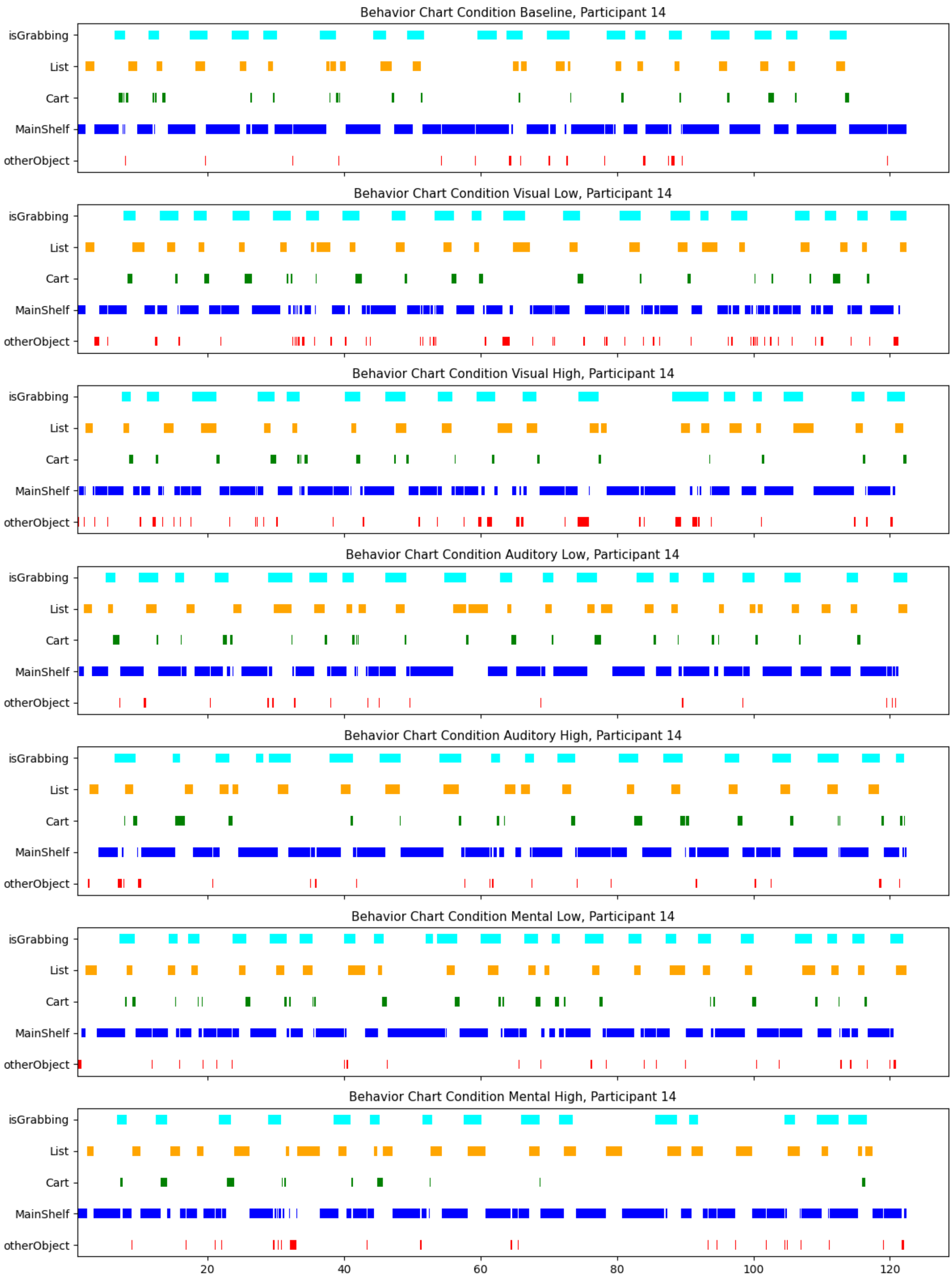


Figure A.25: The behavioral chart illustrates the gaze patterns and object interaction of Participant 14 across all conditions as detailed in the scientific paper. The vertical axis shows the areas of interest within the environment, while colored bars denote the participant's gaze allocation during each condition. Additionally, the presence of object grabbing is displayed. The horizontal axis represents the time, ranging from 0 to 122 seconds.

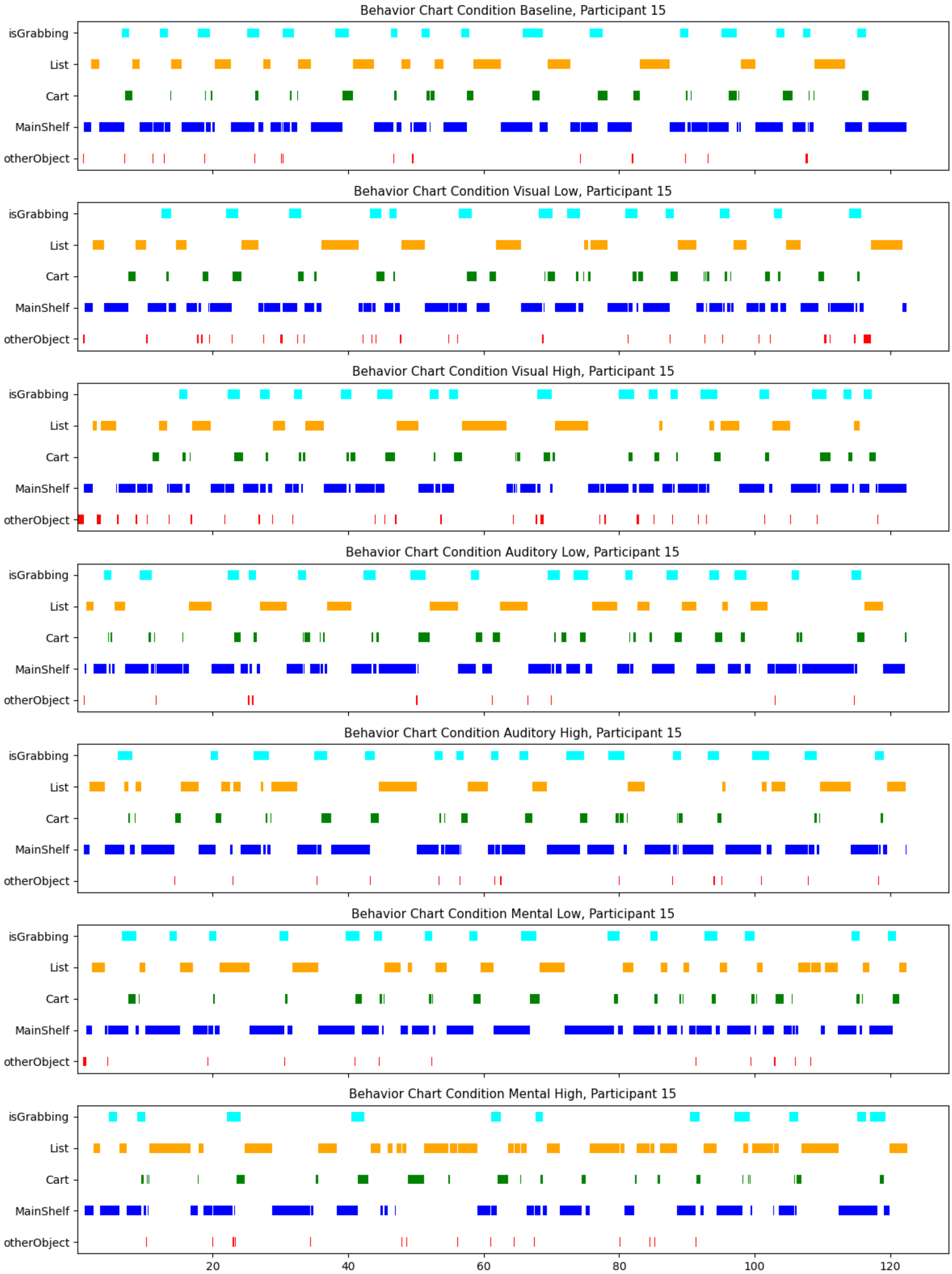


Figure A.26: The behavioral chart illustrates the gaze patterns and object interaction of Participant 15 across all conditions as detailed in the scientific paper. The vertical axis shows the areas of interest within the environment, while colored bars denote the participant's gaze allocation during each condition. Additionally, the presence of object grabbing is displayed. The horizontal axis represents the time, ranging from 0 to 122 seconds.

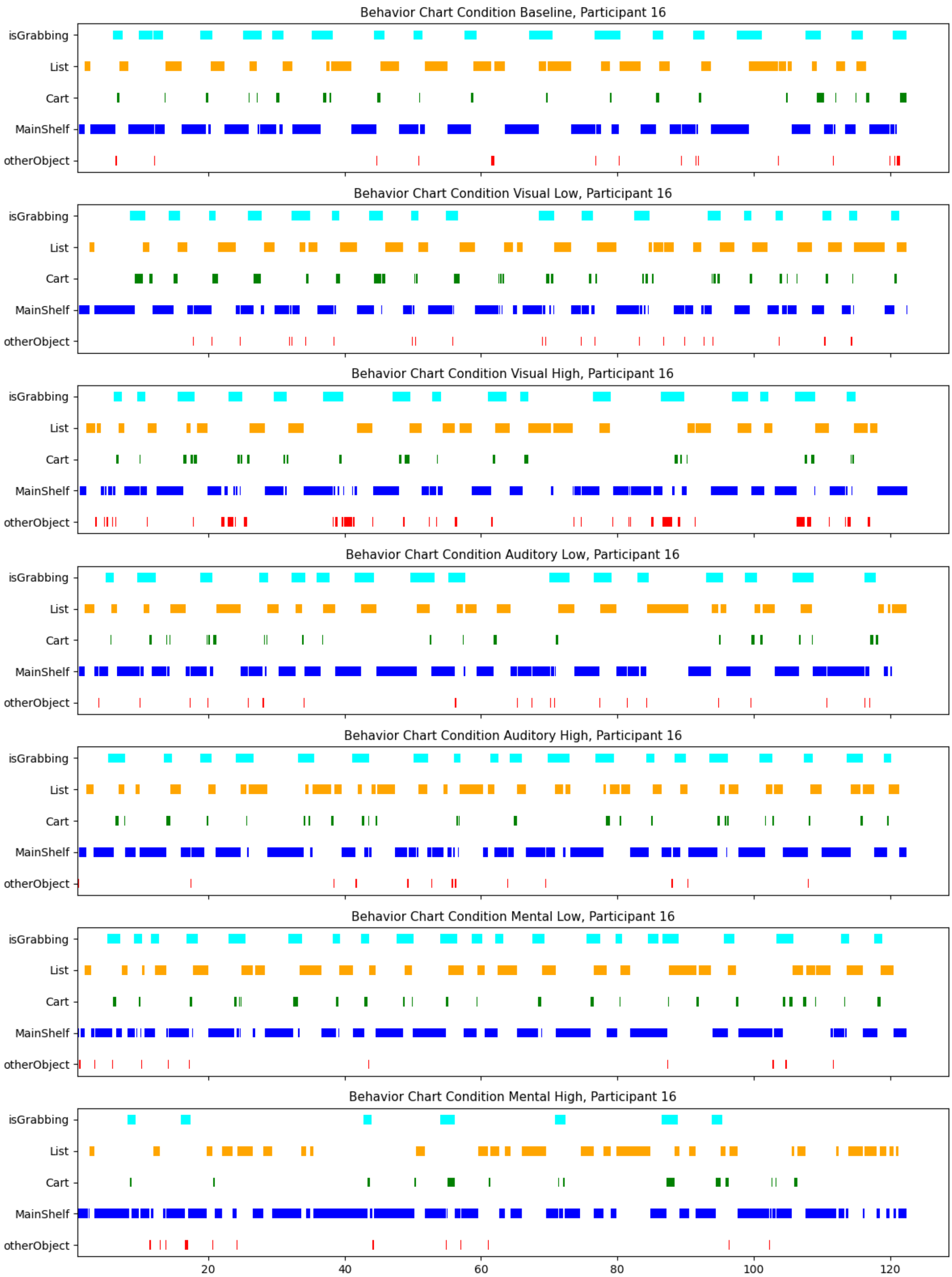


Figure A.27: The behavioral chart illustrates the gaze patterns and object interaction of Participant 16 across all conditions as detailed in the scientific paper. The vertical axis shows the areas of interest within the environment, while colored bars denote the participant's gaze allocation during each condition. Additionally, the presence of object grabbing is displayed. The horizontal axis represents the time, ranging from 0 to 122 seconds.

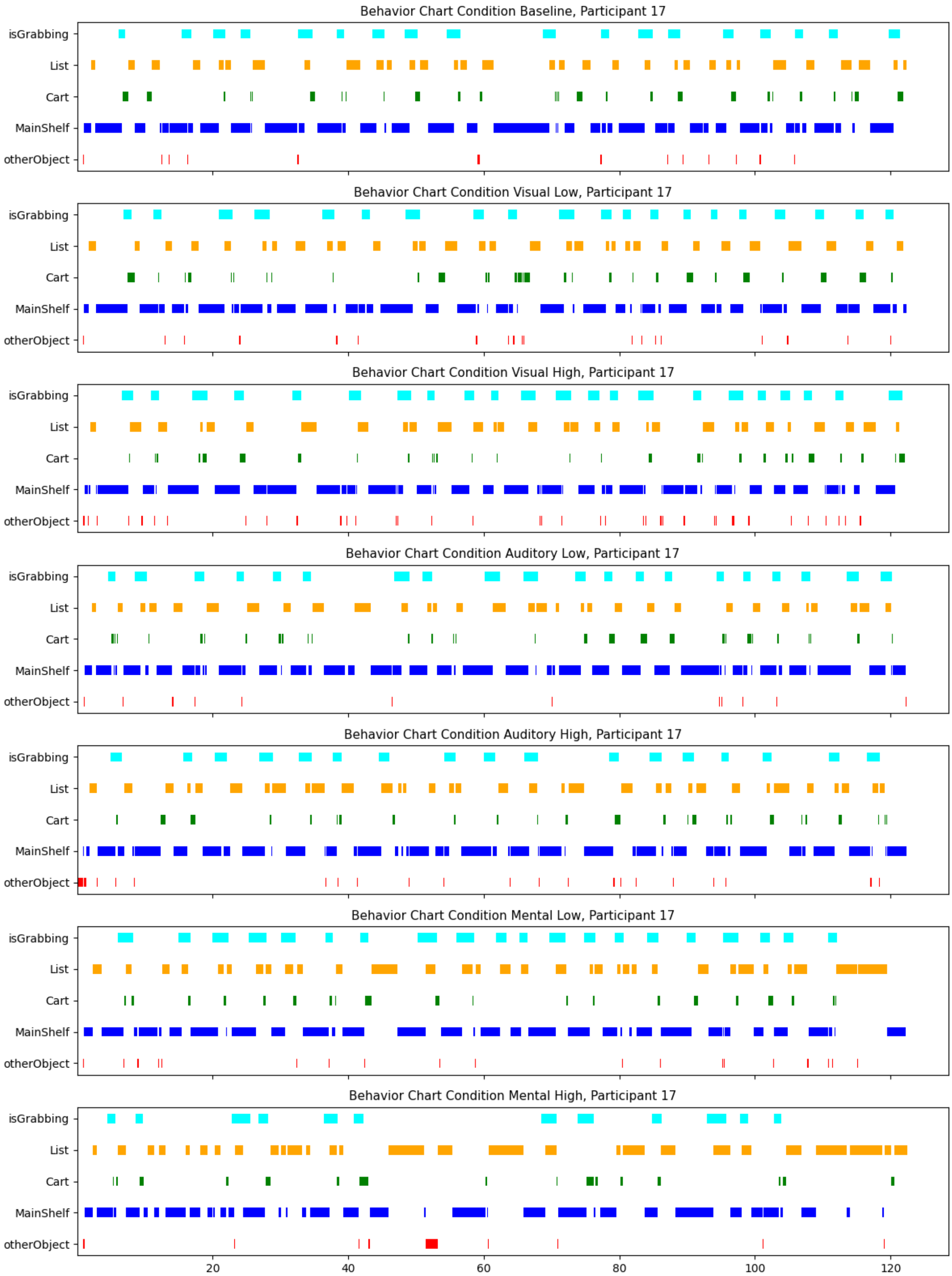


Figure A.28: The behavioral chart illustrates the gaze patterns and object interaction of Participant 17 across all conditions as detailed in the scientific paper. The vertical axis shows the areas of interest within the environment, while colored bars denote the participant's gaze allocation during each condition. Additionally, the presence of object grabbing is displayed. The horizontal axis represents the time, ranging from 0 to 122 seconds.

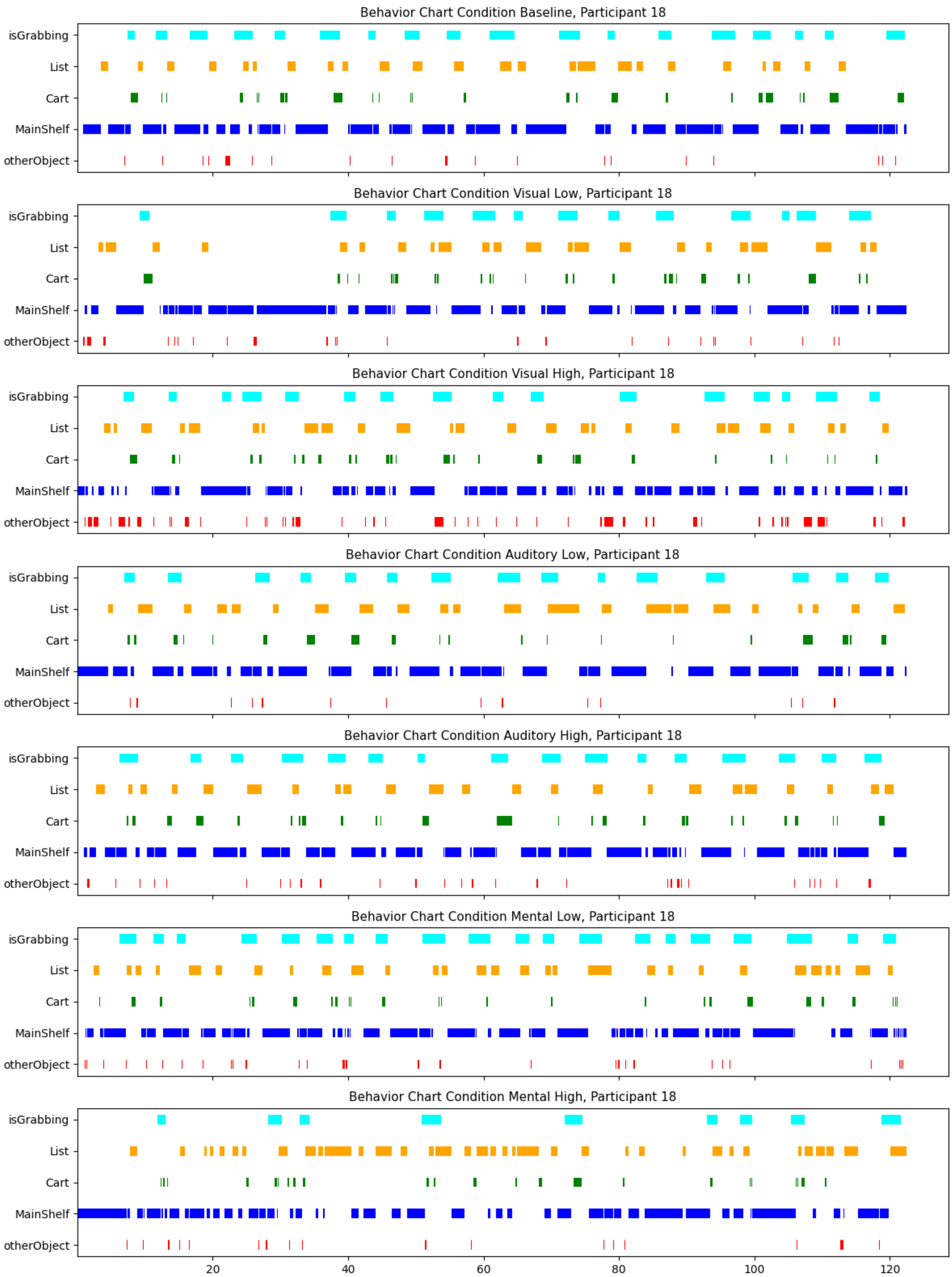


Figure A.29: The behavioral chart illustrates the gaze patterns and object interaction of Participant 18 across all conditions as detailed in the scientific paper. The vertical axis shows the areas of interest within the environment, while colored bars denote the participant's gaze allocation during each condition. Additionally, the presence of object grabbing is displayed. The horizontal axis represents the time, ranging from 0 to 122 seconds.

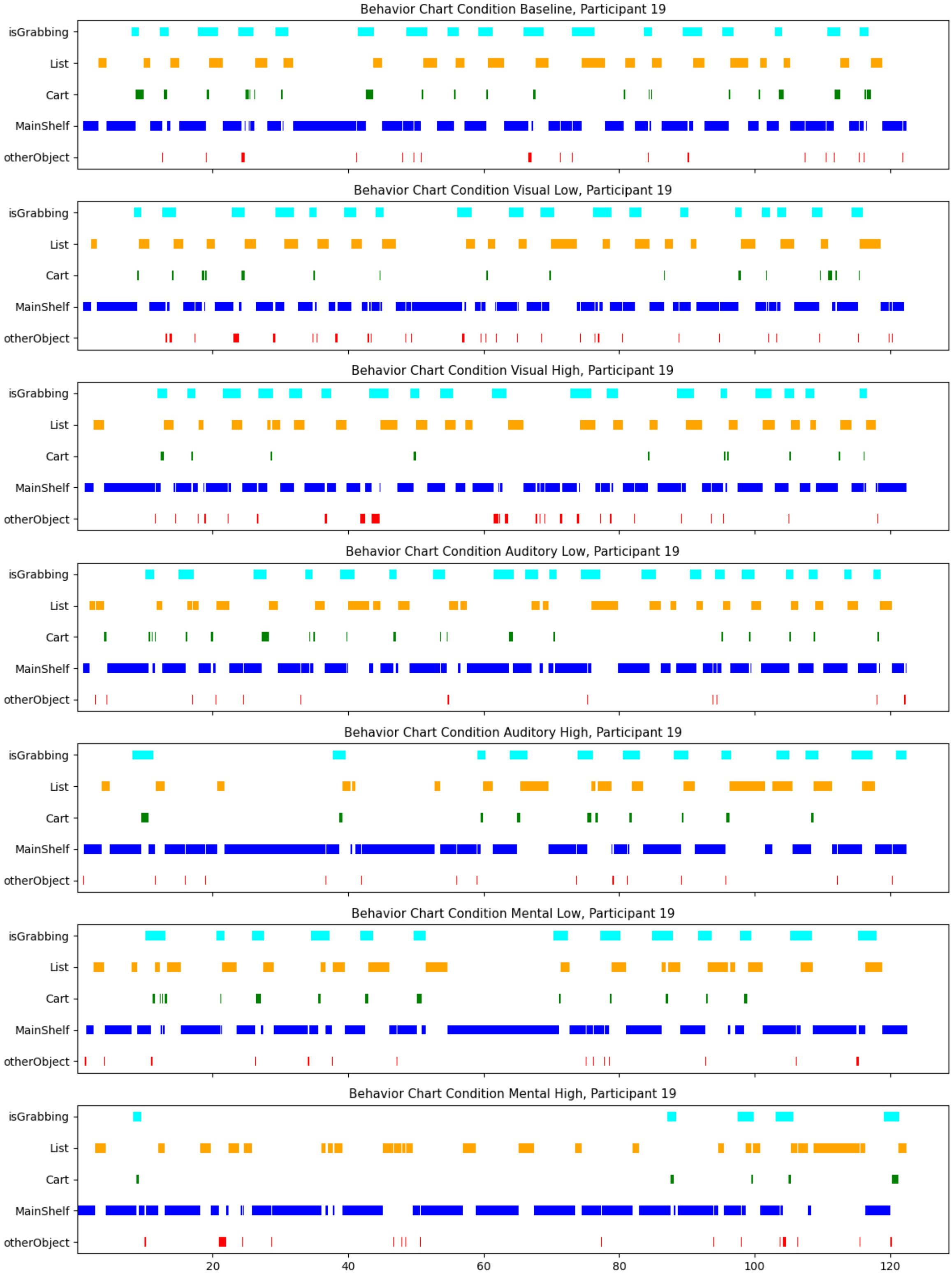


Figure A.30: The behavioral chart illustrates the gaze patterns and object interaction of Participant 19 across all conditions as detailed in the scientific paper. The vertical axis shows the areas of interest within the environment, while colored bars denote the participant's gaze allocation during each condition. Additionally, the presence of object grabbing is displayed. The horizontal axis represents the time, ranging from 0 to 122 seconds.

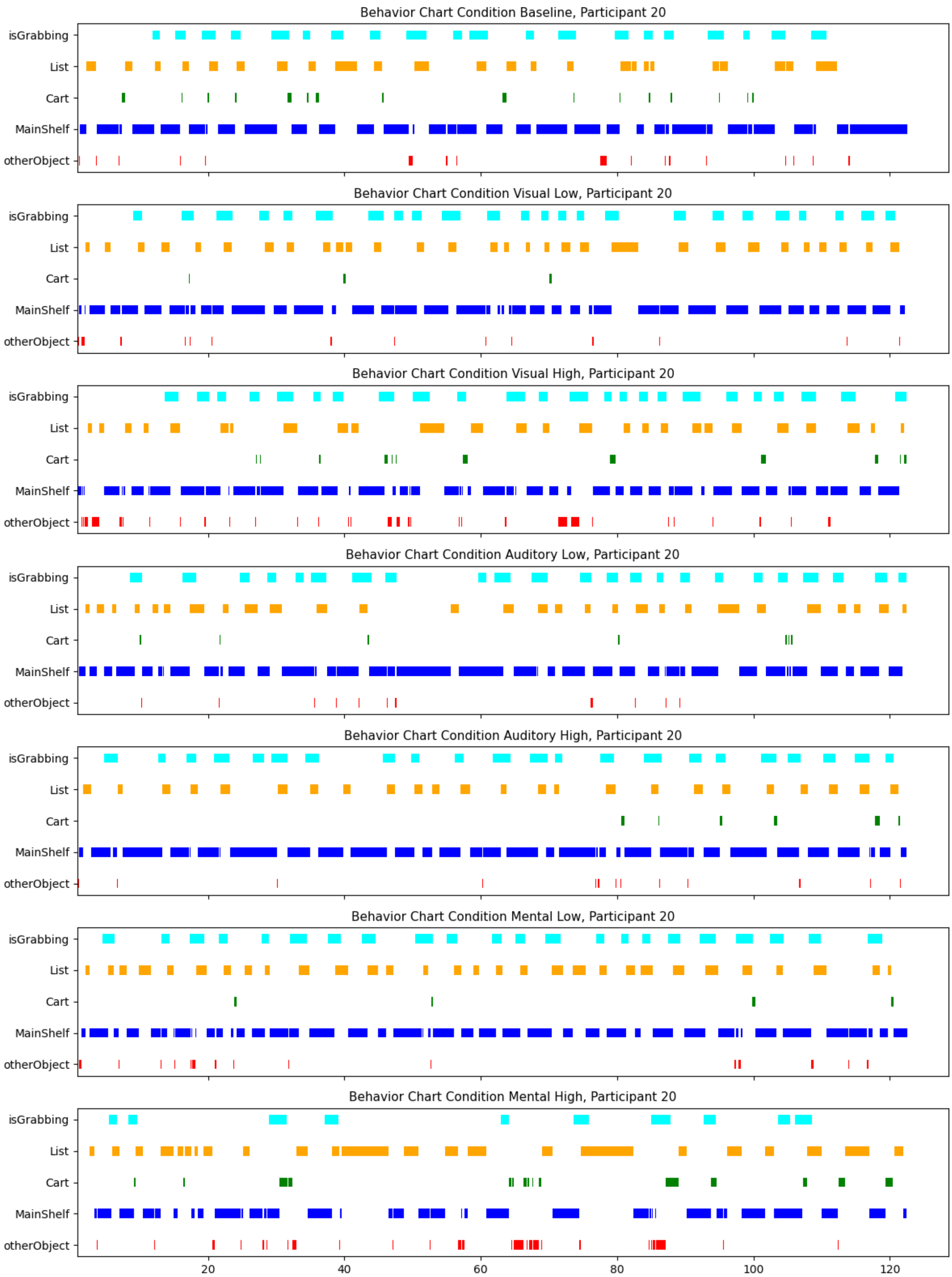


Figure A.31: The behavioral chart illustrates the gaze patterns and object interaction of Participant 20 across all conditions as detailed in the scientific paper. The vertical axis shows the areas of interest within the environment, while colored bars denote the participant's gaze allocation during each condition. Additionally, the presence of object grabbing is displayed. The horizontal axis represents the time, ranging from 0 to 122 seconds.

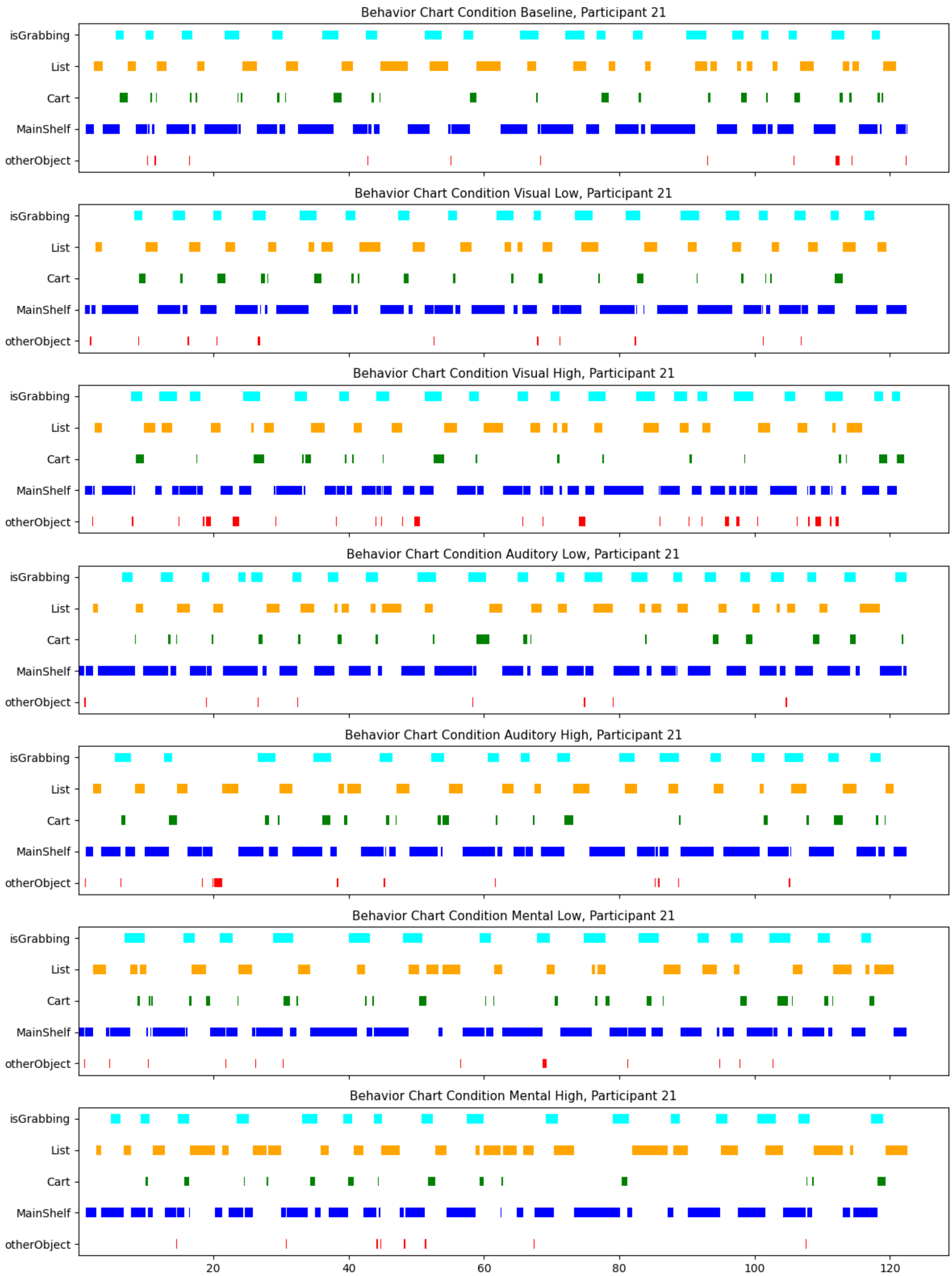


Figure A.32: The behavioral chart illustrates the gaze patterns and object interaction of Participant 21 across all conditions as detailed in the scientific paper. The vertical axis shows the areas of interest within the environment, while colored bars denote the participant's gaze allocation during each condition. Additionally, the presence of object grabbing is displayed. The horizontal axis represents the time, ranging from 0 to 122 seconds.

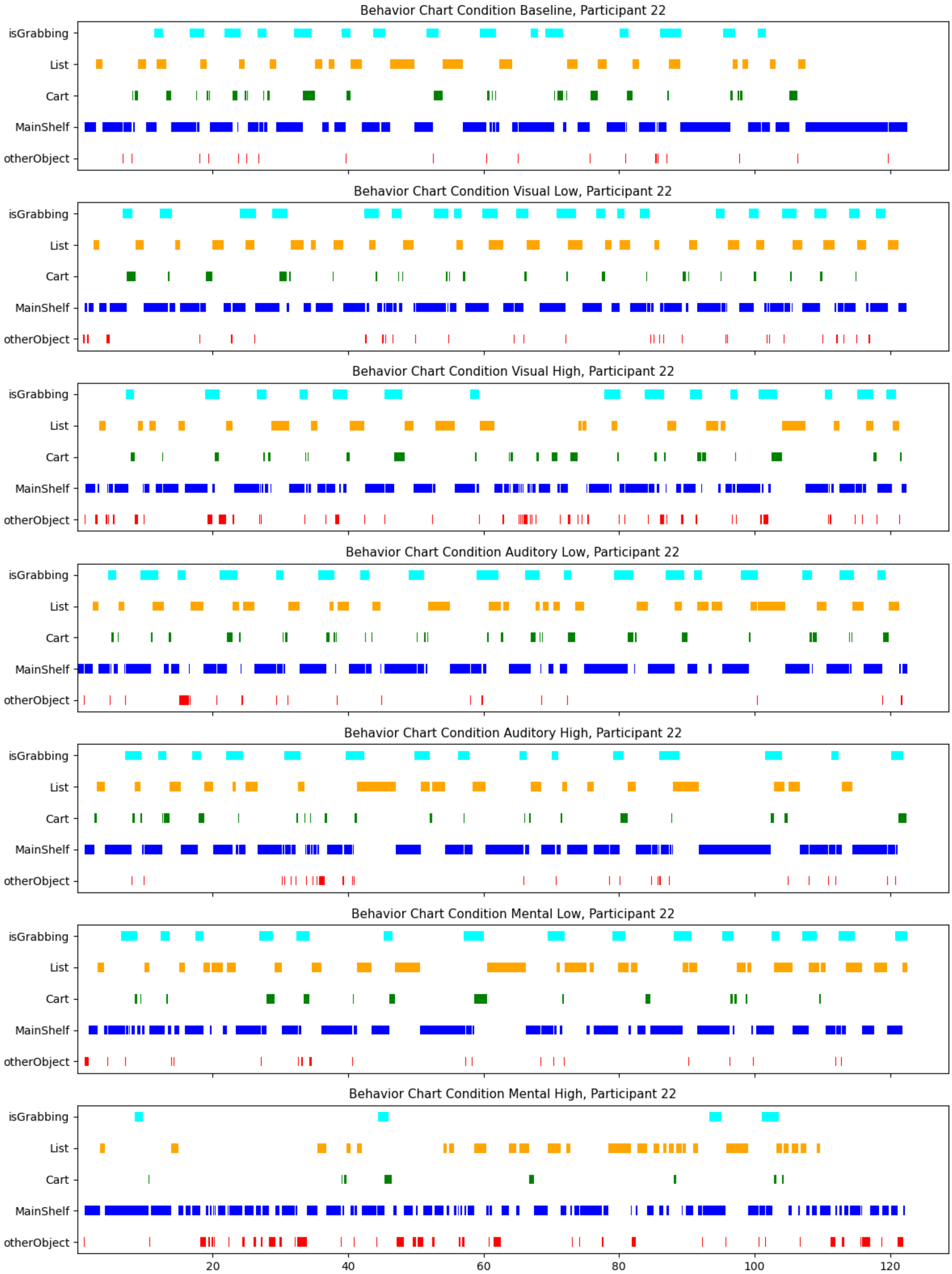


Figure A.33: The behavioral chart illustrates the gaze patterns and object interaction of Participant 22 across all conditions as detailed in the scientific paper. The vertical axis shows the areas of interest within the environment, while colored bars denote the participant's gaze allocation during each condition. Additionally, the presence of object grabbing is displayed. The horizontal axis represents the time, ranging from 0 to 122 seconds.

B

Extensive Environment Description

This appendix gives a more extensive description of the environment in the different conditions. In the first section, the rationale behind the structure of the stimuli is explored. The second section offers an overview of how the rationale of the first section was implemented in the environment, by illustrating how visual objects or audio elements were in motion, the timing of events, and providing additional explanations.

B.1. Design Choices

To organize the stimuli for inclusion in the visual, auditory and mental tasks, hypotheses were formulated regarding their impact on visual and auditory cognitive load. Visual stimuli consist mainly of dynamic character-shaped objects. Auditory stimuli include dynamic or static 3D audio sources, also known as spatial sound. This means that the audio source originates from a specific location, and rotating the head-mounted display towards the source results in corresponding changes in audio sound based on head rotation. The mental task only consists of the audio of the n-back task in stereo, i.e. not spatial.

B.1.1. Visual Conditions

Structure of stimuli in the visual conditions

Factor	Visual Low Condition	Visual High Condition
Character type	Casual characters	Odd, remarkable characters
Character actions	Mainly walking and looking at the participant	Also doing more unique moves, such as waving, dancing, etc.
Events	Only characters	Also non-character objects as visual distractors, such as a flashing alarm.
Location	Behind the main shelf, at least four meters from the participant.	Also next to the main shelf, at least two meters from the participant.
Examples	Casually dressed NPC walking, Casually dressed NPC waving, NPC looking at you	Oddly dressed NPCs such as clown, fireman, farmer, Santa Claus, pilot, Alarm flashing, NPCs dancing, jumping, calling

Table B.1: Overview of the different factors that may influence cognitive load described in this section for the visual low and high condition. Furthermore, examples are given to elaborate more on the factors described in the table.

Explanation table B.1

In the visual condition of the supermarket environment, the main differences will revolve around the types of characters, their actions, the occurrence of events, and the locations of these events/characters. The underlying idea is that, in the low visual condition, subjects may experience the supermarket how you could expect a supermarket to be with a few visuals that can catch the eye, while in the high visual and auditory condition, the stimuli will significantly interact with and distract the subject.

During the low visual condition, characters that wander around in the supermarket are casually dressed, so the environment is not different from a normal grocery shop visit. Conversely, in the high visual condition, characters appear who are not expected in a typical supermarket setting. This deviation from the norm is in line with the principles of the Orienting Response theory [8], which explains the automatic reactions triggered by novel or significant stimuli in the environment. In the context of this study, the high visual condition is expected to induce a stronger orienting response due to the presence of unexpected stimuli.

The same principle applies to the actions of the characters. Based on observations in [9], where NPCs did not engage in any unusual behavior, the subjects may have filtered out these NPCs. Therefore, in the low visual condition, characters will sometimes make an action towards the subject making eye contact to capture their attention. Eye contact is specifically important since is has been shown that eye contact is a direct distractor of humans and even increases cognitive load [10]. In the high cognitive load condition, characters will perform more unlikely and attention-grabbing movements, such as calling, dancing, waving, etc.

In addition to characters interacting with the subject, other events may also attract attention and increase cognitive load. Events like flashing alarms in the background may capture the subject's attention since the orienting response will be higher. It is hypothesized that these events will contribute to an increased cognitive load.

Finally, the positioning of characters and events varies between scenarios with low and high visual conditions. Visual cluttering, characterized by the presence of disorganized visual elements within a specific space, is influenced significantly by high proximity [11]. In the low visual condition, characters will mainly be situated behind the main shelf. Conversely, in the high cognitive load condition, visual stimuli will be situated across various locations, primarily closer to the subject and next to the main shelf. This arrangement aims to intensify both proximity and visual cluttering, thereby potentially increasing cognitive load.

B.1.2. Auditory Conditions

Explanation table B.2

The primary distinction between the low and high auditory conditions lies in the level of distraction, encompassing dynamics, volume, duration, and location. These factors will be discussed in this subsection.

Audio dynamics pertains to whether the audio source remains static or moves. Although there is no overarching theory specifically dedicated to the relationship between dynamic and static sounds and distraction, dynamic sounds have the potential to be more unexpected. The Attentional Control Theory [12], refers to the ability of an individual to select which stimuli to focus on and which ones to disregard, and can be "captured" by salient or unexpected stimuli. In the case of the low auditory condition, the audio is expected to be more static. Since it serves as a background noise, it is less likely to be a significant distraction compared to the high auditory condition. However, to avoid monotony, certain audio sources in this condition may have slow movements. In the high auditory condition, the audio is much more dynamic. For instance, quick footsteps approaching from behind or conversations while in motion may be more distracting.

Volume is heavily influenced by the proximity of the audio source to the subject. When the source is closer, the volume will be higher. According to cognitive load theory, the volume contributes to an increase of extraneous cognitive load [13]. Therefore, in the low auditory condition, the volume will be slightly lower compared to the high auditory condition. Additionally, sounds with short duration (such as objects falling) will be louder than sounds with a longer duration (such as conversations).

Structure of stimuli in auditory conditions

Factor	Low Auditory Condition	High Auditory Condition
Dynamics	Static or translating at a walking pace.	Static or translating at a walking, running, or driving pace.
Volume	Approximately 85 dB.	Approximately 95 dB.
Duration	Continuous audio fragments.	Continuous and short, repeating audio fragments.
Location	At least five meters from the starting point of the participant	At least two meters from the starting point of the participant.
Examples	Checkout desk bleeps, Couple having an argument, Busy crowd Walking footsteps behind main shelf	Police sirens coming by, Person calling to subject nearby, Running footsteps behind subject, Truck in reverse, Bottle falling nearby

Table B.2: Overview of the different factors that may influence cognitive load described in this section for the auditory low and high condition. Furthermore, examples are given to elaborate more on the factors described in the table.

In the low auditory condition, audible audio fragments will vary in duration, ranging from 10 to 300 seconds, with a predominant presence of longer fragments. The Cocktail Party Effect refers to the brain's ability to focus on a specific stimulus while filtering out other monotonous stimuli, usually, auditory [14]. Consequently, subjects are likely to respond more promptly to brief audio fragments, such as the sound of a falling bottle nearby, compared to extended conversations where the auditory information may be more easily filtered. The high auditory condition also includes longer fragments but incorporates shorter, more intense ones as well.

Regarding the location of the audio, in the low auditory condition, the majority of the audio (approximately 75%) will not be near the subject, i.e. at least five meters away from the starting position of the subject. For the high auditory condition, audio sounds can be closer, i.e. at least two meters away from the starting point of the subject.

The audio fragments in the different conditions will aim to be realistic. In the low auditory condition, audio elements will include sounds typically heard in a supermarket environment, such as the audio from the checkout desk with a bleeping device, background conversations among the shoppers, footsteps on various trajectories within the supermarket, and occasional conversations from individuals in the background (one at a time).

For the high auditory condition, the number of conversations will increase, simulating a more crowded and bustling environment. Additional audio elements will be introduced, such as police sirens outside the building coming by, footsteps becoming more intense (moving faster and closer), and a variety of short, louder sounds. These louder sounds may include the effect of bottles falling or a person calling the subject from a close distance.

B.1.3. Mental Conditions

The mental conditions contain the audio of the n-back task, specifically the 0-back and the 2-back task. The n-back task was chosen since it has proven to increase cognitive load for subjects, with more cognitive load [15]. Hence, to create different levels for the mental stimuli, the 0-back (easiest version of the n-back task) and the 2-back task (cognitively harder version of the n-back task) were used in the experiment. By choosing the n-back task as the secondary task, it was ascertained that the low and high mental conditions yield different results.

B.2. Environment Overview

To give a clearer overview of the layout of the environment with visual, auditory, and mental stimuli, activity maps of the environment were created. The map of the baseline condition is not added in this section, since no extra stimuli are displayed on the map. The baseline condition would consist of a map without any actions. Since the visual and auditory conditions contain many more actions, this section goes much more into detail for the low and high visual and auditory tasks.

All maps are presented in Figures B.1, B.2, B.5, B.6, and B.7. A blank environment map (the baseline) consists of the top view of the supermarket environment, including a time axis ranging from 0 to 120 seconds. The colorless rectangles represent product shelves, with the one labeled "Main Shelf" representing the primary product shelf where participants collect items. Colorless circles represent pillars in the environment, the cart is depicted by a rectangle labeled "cart," and the red cross signifies the starting position of the subject. The colored items denote added audio sounds or visuals in the conditions, with color having no specific meaning.

B.2.1. Visual Conditions

The two visual maps (Figures B.1 and B.2) employ different colors to denote events on the time axis. Green diamonds represent idle characters (characters remaining stationary throughout the entire condition), with arrows indicating the direction they are facing. Colored lines on the maps illustrate trajectories along which characters move. Small circles at the ends of the lines indicate the starting and ending points of the trajectory. A trajectory can be traversed in two directions, leading to two distinct trajectory names (e.g., T2 and T3 share the same trajectory in Figure B.2 but in opposite directions). It's important to note that T6 and T7 in Figure B.2 are trajectories where the character moves toward a point, performs an action, and then returns to the starting point. If a diamond is positioned on the line, it signifies that characters on that trajectory are executing an action, as specified on the time axis at that location on the trajectory.

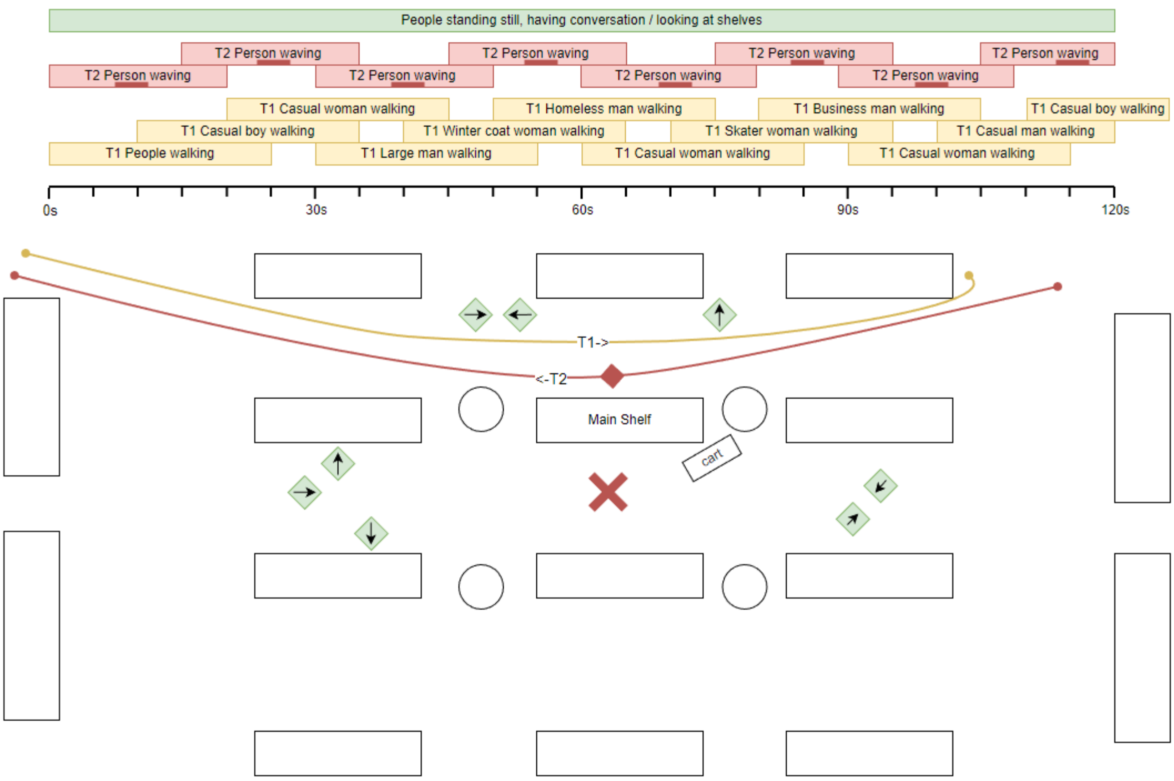


Figure B.1: Action map for the visual low condition. Green diamonds denote stationary characters, with arrows showing their direction. Colored lines represent character trajectories, with small circles marking start and end points.

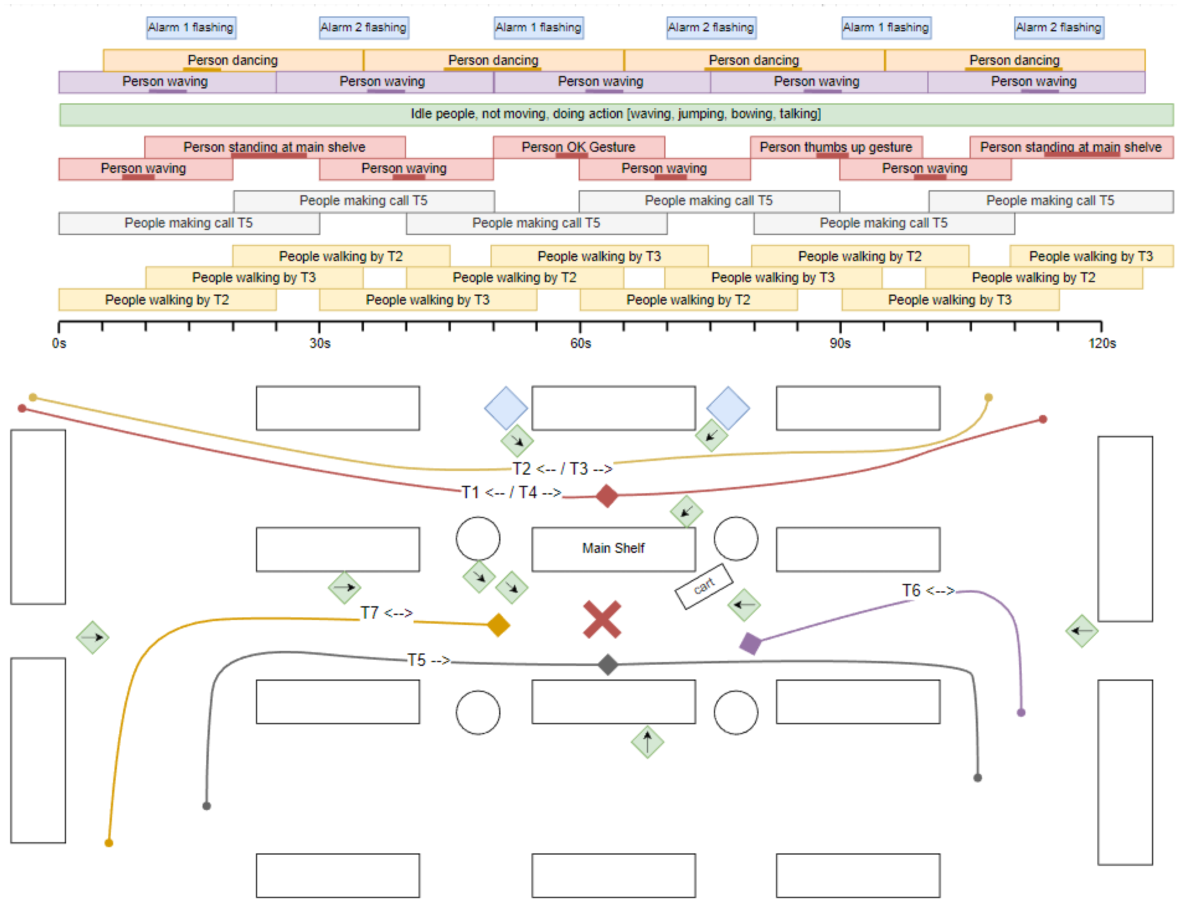


Figure B.2: Action map for the visual high condition. Green diamonds denote stationary characters, with arrows showing their direction. Colored lines represent character trajectories, with small circles marking start and end points. In addition to the visual low condition, this condition contains two flashing lights, displayed with a blue diamond.

Characters appear in the environment at the time allocated in the time schedule and are being destroyed upon completing the trajectory. This design choice prevents the environment from becoming overcrowded, avoiding character collisions, aisle congestion, and product displacements. Characters performing actions first walk toward the location where the action occurs. The colored bar within the colored block on the schedule represents the duration when the action takes place. In the visual low condition, the visual high environment map also features two blue diamonds representing flashing lights, which activate at designated times. In contrast to the characters, the flashing lights neither spawn nor disappear.

The paper describes two types of characters, characters that are casually dressed and characters that are oddly or remarkably dressed. Casually dressed characters are characters that you could expect when walking through the supermarket. Some examples of those characters can be found in figure B.3.

On the other hand, oddly or remarkably dressed characters are the ones that you would not expect in a supermarket. These characters are often dressed in clothes they wear for work, to catch the attention of the subject. Examples of odd characters (both in man and female versions) that are used in the visual high condition are an action hero, sumo wrestler, soldier, paramedic, reporter, metalhead, swat team member, tennis player, butler, post-man, explorer, mechanic, ninja, police, clown, Eskimo, naval officer, farmer, Santa, Viking, carpenter, pilot, and explorer. Some examples of those characters can be found in figure B.4.

Figure B.3: Illustrations of ”casual” characters in the visual low condition. Arranged from left to right, top to bottom: large man, casual boy, casual woman 1, skater girl, casual woman 2, casual winter girl, business man, casual man.

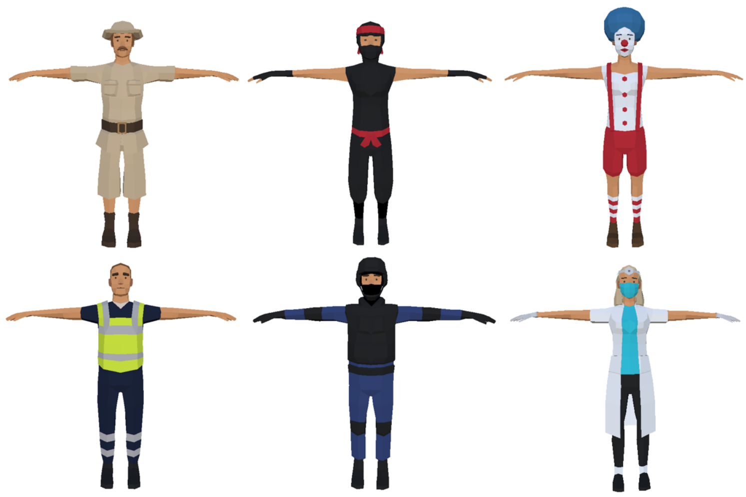


Figure B.4: Illustrations of ”odd” characters in the visual low condition. Arranged from left to right, top to bottom: explorer man, ninja man, clown woman, paramedic man, swat officer man, dentist woman.

B.2.2. Auditory Conditions

As mentioned before, audio sounds in the auditory conditions are spatial. To achieve this in the Unity environment, the Resonance Audio Listener plugin was used. The sound field represents full 360° spatial audio by encoding sound waves on a virtual sphere around a listener. The spatial feature makes it possible to for example locate the checkout cassier bleeps at a specific location in the supermarket, improving the realism of the environment. Similar to the visual condition, the colored diamonds are idle sounds (sounds that remain in the same position throughout the condition). The description of the idle sounds can be found in the schedule. The lines represent trajectories where footsteps are translating. Different types of footsteps could be heard, also translating on different velocities along the trajectory. In the schedule there are three lengths for the blue blocks; (1) the shortest are running footsteps on high heels, (2) the middle length blocks are shoes inside the building, walking at normal speed and (3) the longest blocks are heavyweight boots walking slowly. Sounds were looping until they reached the end of the trajectory, where they were destroyed. In the auditory high condition, the purple diamonds represent incidental sounds, which are sounds that are short and occur only once. The incidental sounds are; (A1) a bottle falling and breaking on the ground, followed by a male voice saying ”I can’t believe you’ve done this!”, (A2) a female voice yelling ”What are you doing?!” and (A3) a male voice screaming ”Hey!”.

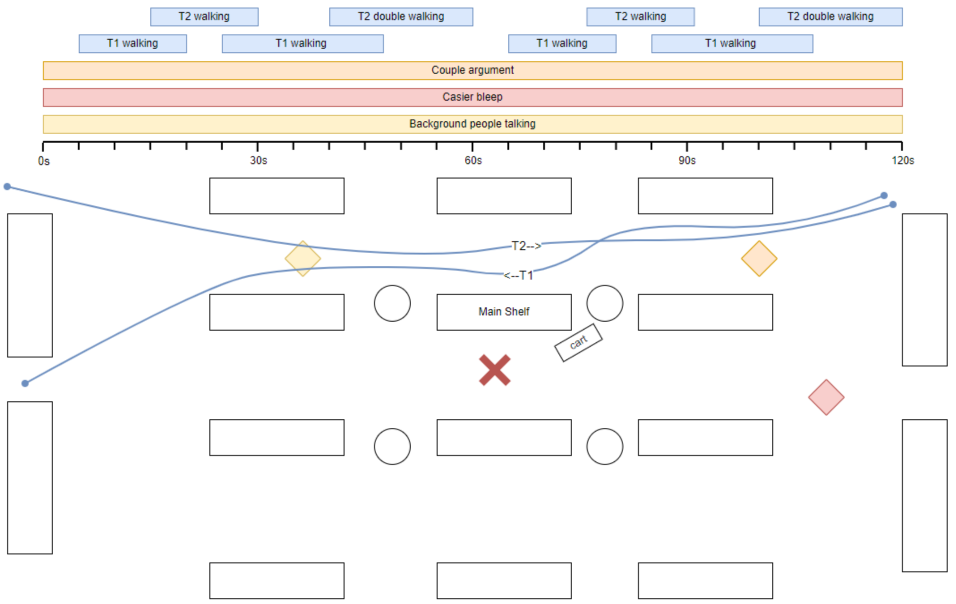


Figure B.5: Action map for auditory low condition. Colored diamonds indicate locations of audio described in the timeline. Two blue lines represent trajectories T1 and T2, depicting footsteps moving in opposite directions.

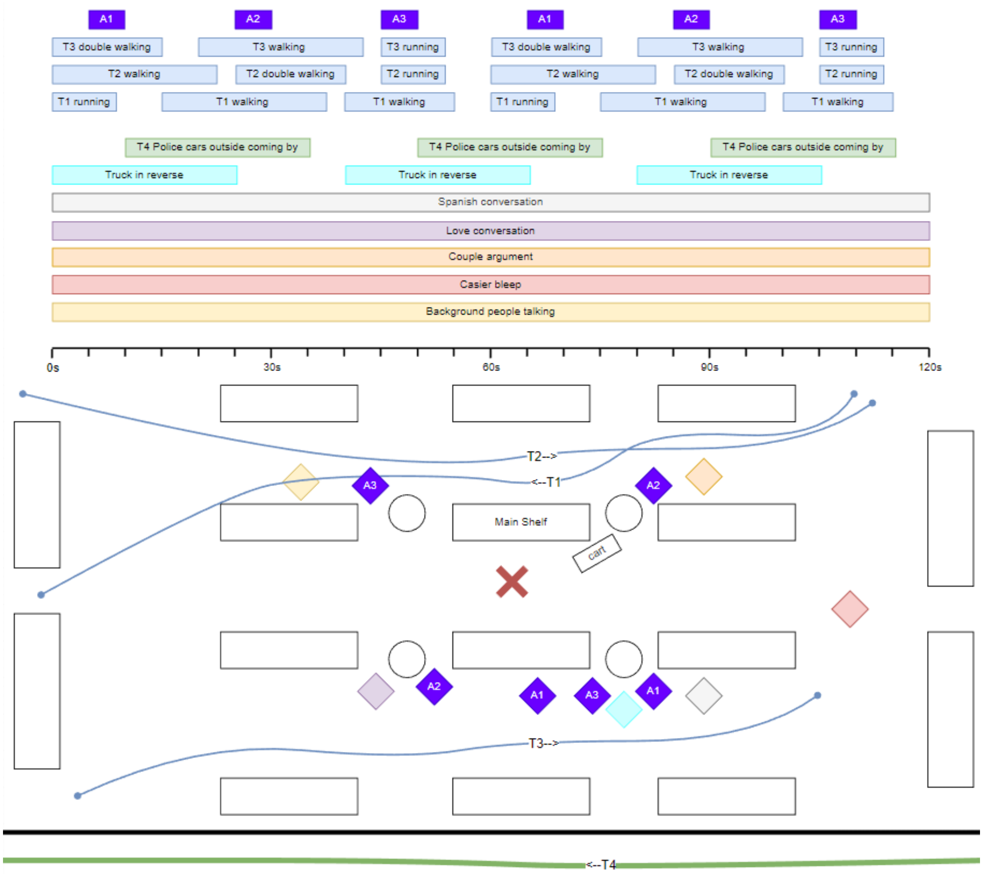


Figure B.6: Action map for auditory high condition. Colored diamonds mark locations of audio described in the timeline. Three blue lines represent trajectories T1, T2, and T3, depicting footsteps moving in opposite directions. Green trajectory T4 depicts the trajectory of police cars passing by outside the building. Dark purple diamonds represent incidental audio sounds, as described in section B.2.2.

B.2.3. Mental Conditions

Since the mental conditions only contain the n-back task, the action map only consists of this action. Contrary to the auditory conditions the audio is played in stereo, which means the audio has an equal volume distributed over the headphones, independent of the rotation of the head. The n-back tasks start after five seconds so that the subjects have some time to adjust to the environment.

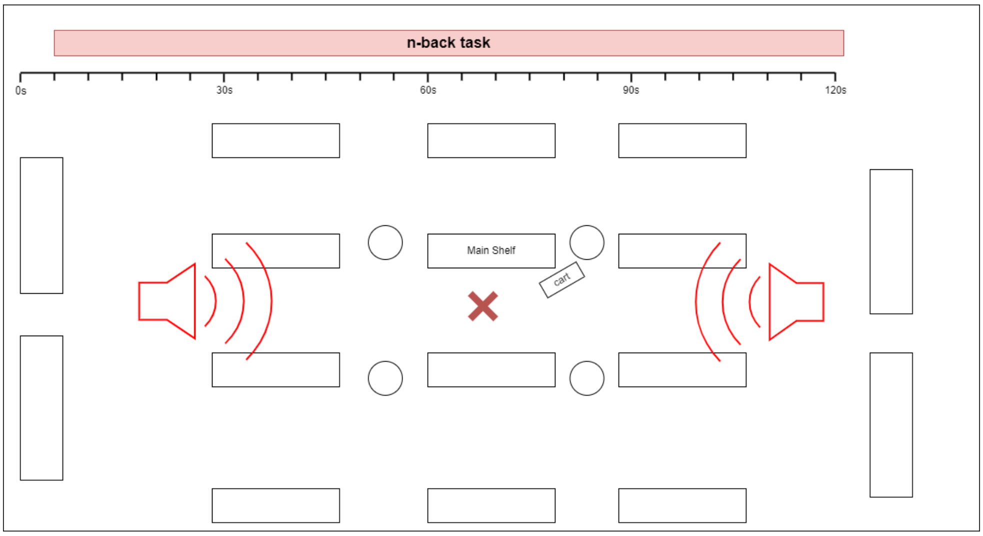


Figure B.7: Action map for mental low and high condition. In these conditions, the audio of the n-back task is played in stereo.

B.3. Unity Design and Scripts

All conditions have their scripts to run the environment. This section describes the most important scripts for the conditions, and how they are used in Unity. The code can be found in this Github repository.

B.3.1. General Scripts

One of the scripts that is used in all conditions is the data collector, which is called *DataCollector.cs*, and the scripts belong to the data collector. The script collects data on head movements, hand movements, and eye-related data. After each condition, data could be saved automatically in a *.csv* file. Similar to all script files, the data collector has some parameters that should be set in the inspector. A screenshot with the data collector script in the inspector can be found in Figure B.8

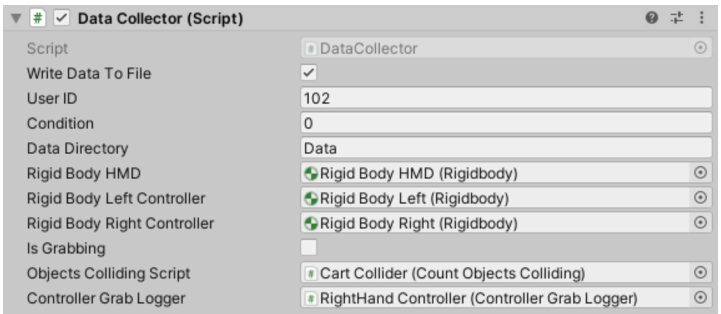


Figure B.8: Caption

Before running the scene in the desired condition, the user ID, condition, and data directory should be modified. Once the condition is finished, the data is then saved in the data directory, directory *P{userID}*, filename *datafile_C{condition}.csv*, but only if the checkbox for ”Write Data To File” is checked. The data file requires the game objects of the HMD and the controller to track position and rotation, so these should be attached to the data collector object. Also, the collider of the cart and the controller that should be logged should be added to the data collector game object in the inspector.

Besides the data collected within the data collector script, it uses three other important data collectors. All collectors update every frame when the condition is running, and collect the data for the data collector script. These three external data collectors are:

- *ControllerGrabLogger.cs*
- *FixationOnObject.cs*
- *CountObjectsColliding.cs*

ControllerGrabLogger.cs The controller logger script collects data about the right-hand controller. When the right-hand controller grabs an object, the isGrabbing parameter is set to true and the name of the object is logged.

FixationOnObject.cs The fixation on object script is attached to all items that should be registered by the TobiiXR package. When a subject focuses on an object with the FixationOnObject.cs script attached to it, the name and tag corresponding to that item are logged. When an object does not contain this script, the name, and tag will be automatically set to ”notAssigned”.

CountObjectsColliding.cs The count objects colliding script uses the collider within the cart and registers when an item is collected. Once an item is added to the cart, the counter is increased by one and the name of the item is added to the list. Moreover, after 2.5 seconds, the item is destroyed since the cart becomes too full.

B.3.2. Scripts Regarding Visuals

The visual conditions contain script to let the characters move along specific trajectories, perform actions, spawn characters and destroy them, not collide them with other characters, and more. Three main scripts in this section are outlined:

- *NavigationNpc.cs*
- *CharacterController.cs*
- *NpcSpawner.cs*

NavigationNpc.cs
The navigation script is used for characters that have to follow one of the trajectories, without acting on the trajectory. In the visual low condition, this script is attached to the characters on T2, in the visual high condition it is attached to T2 and T3. Empty game objects could be created and placed in the designated location. These game objects could be added to the inspector in the NavigationNpc.cs script object. See Figure B.9 for a screenshot of the inspector in Unity.

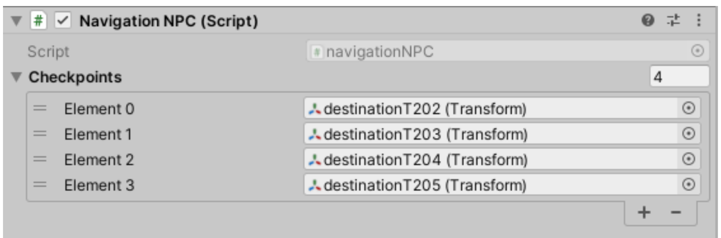


Figure B.9: Display of the attached script of NavigationNpc.cs on an object with its public parameters. The script attached to this character touches four destinations (T202, T203, T204 and T205).

CharacterController.cs
The character controller script works identically to the navigation script but adds an action to the trajectory. In the low condition, this script is attached to the characters on T1, in the visual high condition

it is attached to T1 and T4 to T7. Empty game objects are also added to the inspector. In addition, the number of the checkpoints where the action should be performed, the name of the action, and the duration of the action should be defined. A screenshot of the inspector in Unity for the CharacterController.cs script can be found in Figure B.10.

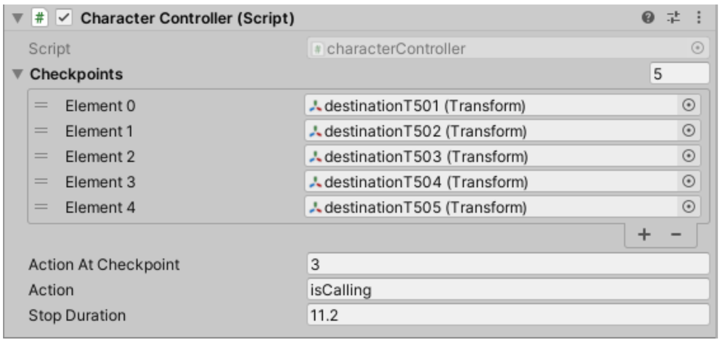


Figure B.10: Display of the attached script of characterController.cs on an object with its public parameters. The script attached to this character touches destination T501 to T505 and performs the "isCalling" action for 11.2 seconds on the third checkpoint (T503).

NpcSpawner.cs
The NPC Spawner is the script that spawns the characters after the desired number of seconds. If the characters are not attached to the array of the NpcSpawner.cs script in the inspector, the characters spawn directly when the condition starts. The character should be added to the "NPC Game Object" with the number of seconds after which the character should spawn. For NpcSpawner.cs script the inspector in Unity is shown in Figure B.11.

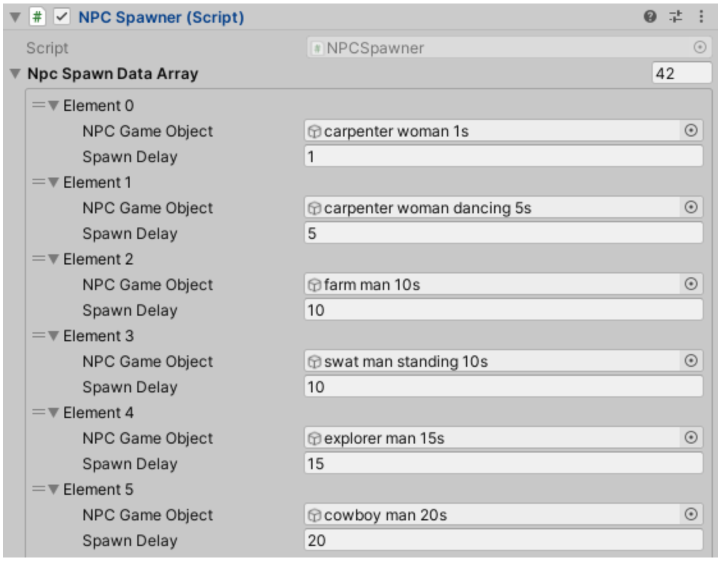


Figure B.11: Display of the attached script of NpcSpawner.cs on an object with its public parameters. The spawner consists of 42 characters that spawn in the visual high condition, with the first character (the carpenter woman) that spawns 1 second after the condition starts.

B.3.3. Scripts Regarding Audio

The auditory conditions were more straightforward to implement as they do not contain animations, colliders, navigation areas, or other difficulties in comparison with the visual conditions. The auditory conditions use the ResonanceAudioMixer in order to spatialize the sounds. Besides, the two most used scripts in the auditory conditions are:

- *AudioTimer.cs*

- *PathFollower.cs*

AudioTimer.cs
The audio timer script is more or less the same as the spawner script in the visual condition scripts. The audio timer starts the audio after a specific number of seconds and loops it for a desired number of times before it stops. The parameters that should be added in the inspector are the audio source, the number of seconds after which the audio should start playing, and the loop count. Refer to Figure B.12 for a screenshot of the inspector. The audio timer script is used in both auditory and mental conditions.

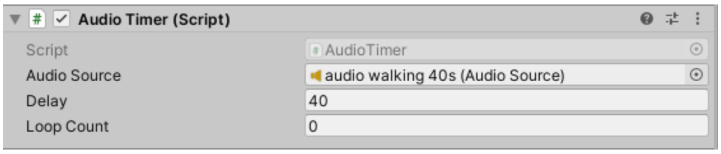


Figure B.12: Display of the attached script of AudioTimer.cs on an object with its public parameters. In this figure, the ”audio walking 40s” is added to as an audio source, which is one of the footstep audios. The footsteps start playing after 40 seconds and do not loop.

PathFollower.cs
The path follower script ensures that the audio source added to the path follower translates over the trajectory at the desired speed. First of all, the trajectory on which the audio should move should be added in the inspector. Second, the end of the path instruction should be given, choosing from ”stop”, ”reverse” and ”loop”. Thirdly, the speed of the audio should be determined. For slow walking footsteps, the speed was chosen to be 1, for running footsteps the value was 3. Lastly, you could choose on which side of the trajectory the audio should start and if the audio should be destroyed after reaching the end of the trajectory. An example could be found in Figure B.13

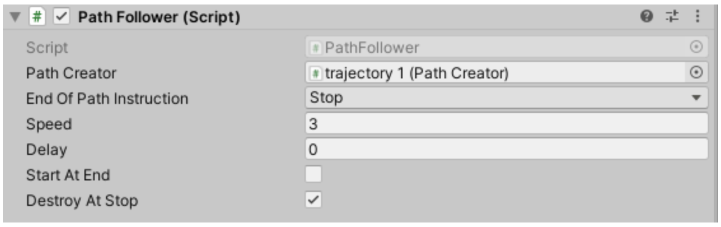


Figure B.13: Display of the attached script of PathFollower.cs on an object with its public parameters. In this case the audio will directly start when starting the condition, it will travel along trajectory 1 with a speed of 3 and it will be stopped and destroyed at the end of the trajectory.

C

R Code Statistical Analysis

To conduct the statistical analysis for this research, R-Studio was employed. As outlined in the scientific paper, a linear mixed-effects model was used for all measures, where the conditions served as fixed effects, the participants as random effects, and the measures as dependent variables. Additionally, the order was added as an effect. This modeling approach was applied to both the predefined and behavioral measures. To control the family-wise error rate, Holm's correction was implemented. The following section presents the R code used for conducting the statistical analysis.

Listing C.1: R Code for Statistical Analysis, Dependent Measures

```
# import libraries
library(lmerTest)
library(Matrix)
library(xtable)

# file path to csv
file_path <- "C:\\path\\to\\datafile\\dataframe_1_R.csv"

# create dataset from csv
dataset <- read.csv(file_path)

# prepare data set
dataset = as.data.frame(dataset)
dataset$condition = factor(dataset$condition, levels = 1:7)
dataset$order = factor(dataset$order, levels = 1:7)
dataset$participant = factor(dataset$participant, levels = 1:22)

# build linear mixed effects models
reg1 = lmer(aoi_cart ~ condition + (1 | participant) + order, data = dataset)
reg2 = lmer(aoi_list ~ condition + (1 | participant) + order, data = dataset)
reg3 = lmer(aoi_main_shelf ~ condition + (1 | participant) + order, data = dataset)
reg4 = lmer(aoi_other_object ~ condition + (1 | participant) + order, data = dataset)
reg5 = lmer(hr ~ condition + (1 | participant) + order, data = dataset)
reg6 = lmer(hrv ~ condition + (1 | participant) + order, data = dataset)
reg7 = lmer(hand_smoothness ~ condition + (1 | participant) + order, data = dataset)
reg8 = lmer(head_stillness ~ condition + (1 | participant) + order, data = dataset)
reg9 = lmer(performance ~ condition + (1 | participant) + order, data = dataset)
reg10 = lmer(nasa_tlx ~ condition + (1 | participant) + order, data = dataset)

# show value summary (optional)
```

```
summary(reg1)

# extract p-values from summary
vec_pval1 = coef(summary(reg1))[2:7,5]
vec_pval2 = coef(summary(reg2))[2:7,5]
vec_pval3 = coef(summary(reg3))[2:7,5]
vec_pval4 = coef(summary(reg4))[2:7,5]
vec_pval5 = coef(summary(reg5))[2:7,5]
vec_pval6 = coef(summary(reg6))[2:7,5]
vec_pval7 = coef(summary(reg7))[2:7,5]
vec_pval8 = coef(summary(reg8))[2:7,5]
vec_pval9 = coef(summary(reg9))[2:7,5]
vec_pval10 = coef(summary(reg10))[2:7,5]

# make list of vectors
list_of_pvalues <- list(vec_pval1, vec_pval2, vec_pval3, vec_pval4, vec_pval5,
                        vec_pval6, vec_pval7, vec_pval8, vec_pval9, vec_pval10)

# Concatenate the vectors into a single vector
all_pvalues <- unlist(list_of_pvalues)

# adjust p-values with Holm correction
adjusted_pval = p.adjust(all_pvalues, method = "holm")
```

Listing C.2: R Code for Statistical Analysis, Behavior

```
library(lmerTest)
library(Matrix)
library(xtable)

# file path to csv
file_path <- "C:\\path\\to\\datafile\\dataframe_2_R.csv"

# create dataset from csv
dataset <- read.csv(file_path)

# prepare data set
dataset$condition = factor(dataset$condition, levels = 1:7)
dataset$order = factor(dataset$order, levels = 1:7)
dataset$participant = factor(dataset$participant, levels = 1:22)

# build linear mixed effects models
reg1 = lmer(overlap_grab_list ~ condition + (1 | participant) + order, data = dataset
reg2 = lmer(ratio_frequency_list_items ~ condition + (1 | participant) + order, data
reg3 = lmer(ratio_time_list_items ~ condition + (1 | participant) + order, data = dat

# extract p-values from summary
vec_pval1 = coef(summary(reg1))[2:7,5]
vec_pval2 = coef(summary(reg2))[2:7,5]
vec_pval3 = coef(summary(reg3))[2:7,5]

# make list of vectors
list_of_pvalues <- list(vec_pval1, vec_pval2, vec_pval3)

# Concatenate the vectors into a single vector
all_pvalues <- unlist(list_of_pvalues)
```

D

Experimental Protocol

Protocol Experiment “Assessing Cognitive Load in Immersive Virtual Reality”.

Before participant comes in

- Check if controllers are charged
- Set up lighthouses
- Calibrate location of HMD
- Start own computer with Qualtrics
- Set up connection heart rate + test device
- Place informed consent on the table, next to the computer

Participants comes in

- Ask participants to go to the toilet
- First let the participant read the informed consent and sign it
- Ask if personal questionnaire has been filled out

Introduction

- The experiment is expected to last around **90 minutes**
- You will perform **7 sessions** including familiarization
- After each session you have to fill out a questionnaire
- You will be in a virtual shopping environment, fulfilling a shopping list by selecting items from the shelves and placing them into a shopping cart
- You're likely not going to finish the entire list, so **please don't approach the task as a race**
- Nonetheless, remember that the primary goal is to pick items from the shelves and put them into the shopping cart.
- All items that are on the list are located on the main shelf. If you cannot find the item, search better. Do **not** search for products on other shelves.
- Some instructions:
    - Keep your left arm along your body since movements cause noise in the sensors.
    - Only use your right arm for grabbing objects.
    - Don’t throw the items, but put them in the cart. If you throw and miss, the item is not being registered, which affects the results in a negative way.
    - Items disappear after some seconds from the cart to avoid piling issues.
    - It is important to complete the list in chronological order.
- Show the familiarization environment to the subject and go through the items to avoid confusion. Especially discuss the following items.
    - Blue foil pack (often confused with Pizza Dough)

    - Fanta (The bottle says Santa instead of Fanta)
    - Leek (Some Dutch participants don't know this word)
- Show that also colors are important, give the blue, green and orange yoghurts as an example.
- You can withdraw from the experiment at any time.
- Do you have any questions at this point?

Participant and device setup

- Give the participant instructions about the ECG device and ask consent to attach the electrodes.
    - Input AB on device, channel B, **not** channel A
    - Black on right, on end of collarbone
    - Red on left, under ribs
    - White (ground) on left, on end of collarbone
- Give the participant instructions about the EDA device
    - Not too tight
    - Left hand, index and middle finger
- Give the participant instructions about the HMD
    - Show the button on the left for calibration
    - Explain how the eye calibration works, and that it will be calibrated after each session.
    - Let the participant try on the HMD with correct sizing.
    - Set the height to **185cm** for each participant.

Before starting the experiment

- Repeat the exercise, ask if the participant has any questions.
- Tell the participant you are going to set up the last things (see below)
- Start the EDA and ECG device.
- Start the windows recorder.
- Change the parameters for the data collector script in the Unity inspector (user ID and condition).

During the experiment

- Check for using right arm
- After each condition
    - Fill out the NASA-TLX
    - Do the HMD calibration again
    - Do the eye calibration again
- Change the condition number
- Ask for motion sickness

- When the condition includes the n-back task, explain it and practice with the dummy n-back audio.
- After the n-back condition, tell the participant that the next condition does or does not include a secondary task.

After the experiment

- Give the gift card to the participant
- Add the data to the repository (backup)
- Charge the controllers
- Clean device

n-back task explanation

We are going to add an extra task to the main task, which is called the 0-back task. You will hear an audio voice, which gives you multiple sets of single digit numbers. In this version you repeat out loud the last number you've heard. For example, if you hear the number 3, you would say three. If you then hear the number 7, you say seven. And if you hear 2, you would say two, and so on. The n-back task and the main task have the same level of priority. Try to be as accurate as you can be.

We are going to add an extra task to the main task, which is called the 2-back task. You will hear an audio voice, which gives you multiple sets of single digit numbers. In this version you repeat out loud the number you've heard 2 numbers ago. For example, if you hear 1, you would say nothing. If you hear 2, you would say nothing. If you hear 3, you would say 1, if you hear 4 you would say 2, and so on. The n-back task and the main task have the same level of priority. Try to be as accurate as you can be.

ECG electrode placement – bovenkant net onder sleutelbeen, onderkant net onder ribbenkast.

Connecting the hardware

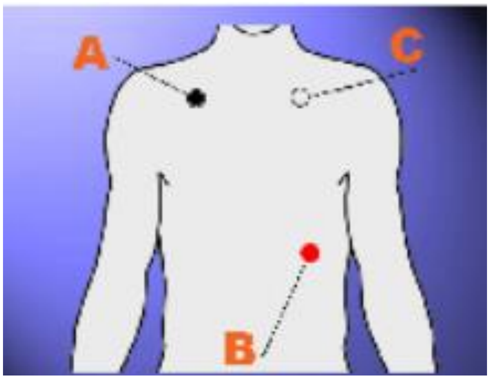
1. Connect the EXG Sensor to input A&B of the NeXus. Connect the Ground Sensor in the left input, labeled Gnd.
2. The default channel for ECG in BioTrace+ uses input B.Thereby, red-2 and black-2 are used for measuring ECG. Snap the appropriate self-adhesive ECG electrodes into the positive red and negative black connectors marked with number 2. See the following overview.

Red-1:	*Not used for ECG by default*	
Black-1:	*Not used for ECG by default*	
Red-2:	Positive input:	Channel B
Black-2:	Negative input:	Channel B

Electrode placement

To acquire an ECG signal that is suitable for heart rate detection, place the negative and positive electrode in such way that the electrical activity of the heart (especially the R-peak) can be measured between the electrodes. Optimally, the electrodes are placed exactly along the electrical heart axis, like the vertical lead II. The positive and negative electrodes should not be switched to avoid incorrect R-peak detection.

The vertical lead II chest placement: Place the red electrode on the lower left rib cage. Place the black electrode on the distal end of the right collarbone. Place the ground electrode on the distal end of the left collarbone. With this electrode placement, the polarity of the R peak is correct.



EDA : wijsvinger en middelvinger

Connecting the hardware

1. Connect the Skin Conductance Sensor to input C of the NeXus.
2. Snap the Ag-AgCL electrodes embedded in Velcro straps into the two connectors.

Electrode placement

Most often, the Skin Conductance electrodes are placed on the fingertips of the nondominant hand, at the palmar side of the fingers. There is no standard that defines which fingers to use. To avoid accidental touching of the electrodes (and thus interference with the signal), one can choose two non-contiguous fingers.

Another option is to place the Skin Conductance electrodes on the palm of the nondominant hand, to create more freedom of movement. In that case, place one special self-adhesive electrode suitable for SC application near the base of the thumb and one near the base of the little finger.


E

Informed Consent

Informed Consent

You are invited to participate in a research study titled "Assessing Cognitive Load in Immersive Virtual Reality." This study is being conducted by Job Mulder, under the supervision of Laura Marchal-Crespo, Joost de Winter and Salvatore Luca Cucinella from the TU Delft.

The experiment is expected to last around one hour and is divided into seven sessions along with a familiarization phase. The duration of each session (excluding the familiarization phase) is fixed: the session concludes automatically after two minutes. Throughout this experiment, you will be immersed in a virtual grocery store environment through a head-mounted display.

Within this setting, your task involves fulfilling a shopping list by selecting items from the shelves and placing them into a shopping cart. It's important to emphasize that the shopping list should be completed in the order provided. You're likely not going to finish the entire list, so please don't approach the task as a race. Nonetheless, remember that the primary goal is to pick items from the shelves and put them into the shopping cart.

After completing the list, we kindly request that you fill out a questionnaire to gather your experiences as a participant. The purpose of this research study is to collect data for analyzing physiological and subjective responses (such as heart rate, skin conductance, eye movements, biometrics, and questionnaires) while performing a task in a virtual environment. These measures are thought to be indicative of cognitive load. In addition to measuring psychophysiological and subjective responses, the screen will be recorded to analyze your behavior during task execution. Audio will also be captured to analyze your responses during the experiment. Finally, photographs may be taken for inclusion in the report, but they will be anonymized before being added.

The collected data will be used to compare responses to different types of stimuli under varying levels of cognitive load. It will also contribute to the development of an algorithm for assessing cognitive load. Furthermore, the data collected from you will be anonymized. Should we decide to publish our research, all sensitive information will be excluded.

We assure you that we will make every effort to maintain the anonymity of your answers in this study. To minimize any risks, we will utilize Project Storage from TU Delft, a secure platform that will protect the collected data.

Please note that during the experiment, you may experience motion sickness or anxiety. Your participation in this study is entirely voluntary, and you are free to withdraw at any time. You may also choose to omit any questions before the experiment begins, although questions will not be answered during the experiment. After the research, if you have any questions or wish to file a complaint, please feel free to contact j.l.a.mulder@student.tudelft.nl.

Herewith I confirm, the undersigned, that I give permission to participate in the study. In connection with this, I declare the following:

	Yes	No
1. Taking part in the Study		
- I have read and understood the study information.	☐	☐
- I have been able to ask questions about the study and they have been answered to my satisfaction.	☐	☐
- I am sufficiently informed about the purpose and procedures of this study.	☐	☐
- I consent voluntarily to participate in this study and I understand that I can refuse to answer any questions and that I can withdraw from the study anytime.	☐	☐
- I am aware that I can revoke myself from the experiment and that the collected data may be used anonymously after 1 week from date signed.	☐	☐
- I understand that taking part in the study involves that information will be captured using ECG sensors, eye-tracking technology, pupillometry, skin conductance sensors, biometric data, audio recordings, pictures, and written notes.	☐	☐
- I agree to have ECG recordings and eyes-movements, pupillometry, skin conductance sensors and biometric data captured while taking part in the study.	☐	☐
- I agree to be photographed while taking part in the study.	☐	☐
- I agree to be audio and screen recorded while taking part in the study to analyse the behavior afterwards, and that audio recordings will be transcribed as text.	☐	☐
- I agree that I will fill out a questionnaire myself.	☐	☐
2. Risks associated with participating in the study		
- I am aware that taking part in this study could evoke motion sickness and anxiety.	☐	☐
3. Use of the information in the study		
- I understand that collected personal information will not be shared beyond the research team.	☐	☐
4. Future (re)use of the information by others		
- I am aware that all the collected data will be anonymized.	☐	☐
- I am aware that the collected data will be stored in the Project Storage at the TU Delft and Git(lab)/subversion repository at TU Delft.	☐	☐
- I give permission for the materials (anonymized) generated in this study to be made publicly available for research in a repository, such as Zenodo.org (EU-Based data repository).	☐	☐
- I give permission to use collected anonymized data to be made publicly available for research.	☐	☐

 Participant № ________

Signatures

For the participant
I have read the information carefully and I agree with the statements.

Full Name	Signature	Date

For the researcher
I hereby confirm that I have fully explained to the participant the purpose and procedure of the study and that they have volunteered to participate in this research.

Full Name	Signature	Date

Bibliography

[1] D. George and M. Mallery. *SPSS for Windows Step by Step: A Simple Guide and Reference, 17.0 update*. 10a. Boston: Pearson, 2010.

[2] J. Hair et al. *Multivariate Data Analysis*. 7th. Upper Saddle River, New Jersey: Pearson Educational International, 2010.

[3] Robert G Pachella and Dennis Fisher. “Hick’s law and the speed-accuracy trade-off in absolute judgment.” In: *Journal of Experimental Psychology* 92.3 (1972), p. 378.

[4] Jens Förster, E Tory Higgins, and Amy Taylor Bianco. “Speed/accuracy decisions in task performance: Built-in trade-off or separate strategic concerns?” In: *Organizational behavior and human decision processes* 90.1 (2003), pp. 148–164.

[5] Philip Sedgwick. “Pearson’s correlation coefficient”. In: *Bmj* 345 (2012).

[6] Giulio Barbato et al. “Features and performance of some outlier detection methods”. In: *Journal of Applied Statistics* 38.10 (2011), pp. 2133–2149.

[7] Nika Goljar et al. “Psychophysiological responses to robot training in different recovery phases after stroke”. In: *2011 IEEE International Conference on Rehabilitation Robotics*. IEEE. 2011, pp. 1–6.

[8] Evgeny N Sokolov et al. *The orienting response in information processing.* Lawrence Erlbaum Associates Publishers, 2002.

[9] Salvatore Luca Cucinella et al. “Shopping in Immersive Virtual Reality: Effects of Diminishing Visual, Auditory, and Cognitive Demands on Workload”. In: (Nov. 2023).

[10] Clara Colombatto, Benjamin Van Buren, and Brian J Scholl. “Intentionally distracting: Working memory is disrupted by the perception of other agents attending to you—even without eye-gaze cues”. In: *Psychonomic bulletin & review* 26 (2019), pp. 951–957.

[11] Daniel V Meegan and Steven P Tipper. “Reaching into cluttered visual environments: Spatial and temporal influences of distracting objects”. In: *Quarterly Journal of Experimental Psychology-Section A-Human Experimental Psychology* 51.2 (1998), pp. 225–250.

[12] Duncan E Astle and Gaia Scerif. “Using developmental cognitive neuroscience to study behavioral and attentional control”. In: *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology* 51.2 (2009), pp. 107–118.

[13] John Sweller. “Element interactivity and intrinsic, extraneous, and germane cognitive load”. In: *Educational psychology review* 22 (2010), pp. 123–138.

[14] Barry Arons. “A review of the cocktail party effect”. In: *Journal of the American Voice I/O society* 12.7 (1992), pp. 35–50.

[15] Susanne M Jaeggi et al. “The concurrent validity of the N-back task as a working memory measure”. In: *Memory* 18.4 (2010), pp. 394–412.