Exploring the Effects of Visual and Auditory Distractors in Virtual Reality on Perceived Cognitive Load and Cognitive Alertness

JACK CLARK, LILA BOZ
College of Information Science,
University of Arizona,
Tucson AZ,
USA

Abstract: - Virtual reality (VR) has been increasingly used within training contexts. It is important to understand how users are affected by various distractions for designing more effective learning experiences in VR. In this paper, we explored how visual (i.e., virtual balls bouncing across the user's field of view) and auditory (i.e., pure-tone click train sounds) distractors affect perceived cognitive load and the relation to cognitive alertness. We conducted a within-subjects user study (N = 48) that revolved around a visuo-spatial cup stacking task. Participants completed seven trial conditions in total (no distractor condition, three visual distractor conditions with varying proximities, and three audio distractor conditions with varying pitches). We measured perceived cognitive load through questionnaires and cognitive alertness through changes in pupil dilation. In summary, participants were negatively affected by the auditory distractors and not the visual distractors.

Key-Words: - Cognitive alertness, cognitive load, distraction, eye tracking, virtual reality, visual and auditory distractors.

Received: June 24, 2024. Revised: January 11, 2025. Accepted: April 5, 2025. Published: May 21, 2025.

1 Introduction

The cognitive resources involved in handling new information are limited and susceptible to getting overloaded through sensory channels, [1], [2]. Through our visual channel alone, we interpret information in distinct modes, such as graphics, written text, movement, depth, spatial awareness of avatars in digital experiences, etc. Each mode can be dissected into fundamental visual cues that trigger visual attention and cognitive recall, [3]. Many aspects of information processing are unconsciously occurring within milliseconds of reception, altering perception, interpretation, and cognitive load on working memory, [4].

Cognitive Load Theory suggests that each mode in working memory has a capacity, [1], [2]. Humans have a finite number of resources available for processing information from each modality (e.g., visual-graphic, text, ambient sound, speech, music, etc.). Also, processing channels can be obscured or delayed when an overabundance of information is being inputted and attempting to be perceived through sensory channels.

The growth of technology has created multimodal environments that may overload information processing capacity which could disrupt focus, [5]. The increased number of modes in multimodal environments like virtual reality (VR) makes it difficult to understand which stimulation channels (e.g., visual, auditory, etc.) are being overloaded at a given time, especially when distracted by extraneous visual and auditory stimuli. Although VR is known to block out interfering information from the outside world due to covering the whole field of view of the user with synthetic imagery, VR experiences can still overwhelm the user's senses through excessive stimuli, [5]. Recent research suggests that a balance of different channels of stimuli is needed for VR to be optimized for learning and training contexts, [5], [6]. VR head-mounted displays (HMDs) and eye-tracking equipment, used in tandem, have been found useful in research to dissect the fundamental cues that affect cognitive load while immersed in multimodal virtual environments, [7]. Understanding these triggers and how they affect information processing can improve learning experiences in VR and in turn, provide enhanced training. Most of the research surrounding this topic has focused on learning outcomes, such as retention, recall, and transfer, [5], [6], [8]. While this research is necessary, it lacks the moment-to-moment focus on the stimuli in training contexts. It is also important to understand the thresholds of our visual and auditory channels so we can avoid distracting stimuli for more effective learning experiences in VR. For instance, what frequency and proximity of extraneous visual and auditory stimulation causes a disruption to focus? Or, at what point does the user make a mistake in a problem-solving activity because there is too much distracting information in the virtual environment? What types of distractions cause the biggest increase in cognitive load (e.g., visual, auditory, etc.)?

The purpose of this research was to measure cognitive performance in VR with the presence of visual and auditory distractors while immersed in a cognitive-motor dual-task paradigm. As suggested in [9], the lack of synthesis across the empirical evidence surrounding cognitive-motor dual-tasks makes it difficult to understand patterns in attention processing that may be occurring regardless of individual differences. There has been some previous research that has begun to highlight some of these patterns for cognitive load and learning outcomes, [10], [11], [12], [13], [14], [15], [16]. However, research on labeling patterns based on distraction type and intensity to see the effects on cognitive alertness is scarce, [9]. In this research, visual and auditory distractions were explored, focusing on cognitive load and alertness when altering pitch, amplitude, speed of motion, and spatial proximity of distractions. These predictors were chosen to contribute to the breadth of empirical evidence available on this topic.

This article is based on the first author's dissertation work, [17]. Another paper from the dissertation work, but focusing on EEG data and user performance can be seen in [18]. Our main contributions in this paper include presenting the results of the user study regarding perceived cognitive load measured through questionnaires and cognitive alertness measured through changes in pupil dilation. We discuss the results in light of our hypotheses and implications for future VR applications regarding more effective learning and training contexts.

2 Related Works

In the early 2000s, researchers began to discover that the amount of focus on attention during high working memory tasks dictates what information is perceived, [1], [2], [19], [20]. For example, when the attention of working memory is on a visual-verbal focal task, the possibility decreases that an irrelevant sound can draw attention, [20]. This signifies working memory has limited capabilities and information is filtered by early sensory attention and reception. It also suggests that irrelevant information is shielded when working memory has increased functioning and/or there is a high task difficulty. Cognitive Load Theory helps explain some of these nuances of working memory and offers a measurable understanding of its mental activity capacity. It surrounds three interacting types of cognitive load: intrinsic, germane, and extraneous, [1], [2]. Intrinsic load is the combination of the number of elements interacting in a task and the individual's previous knowledge. Germane load is dictated by the number of resources needed to accommodate and assimilate the elements of information. The more difficult a task is, the more interacting elements, and in turn, an increase in the resources allocated to the germane load to support the complexity. The extraneous load consists of aspects of a task that cause a mental burden or unnecessary complexity to the task. Often the extraneous load is derived from elements that are not cohesive and/or disruptive to the task.

2.1 Dual-Channel Processing and Split Attention

The threshold for working memory capacity is changing as digital technologies are becoming more widely integrated into everyday routines. The concept of overloading visual and auditory capacity has recently been more heavily researched due to these emerging technologies. The Split Attention Principle states that learning outcomes are decreased when viewer attention is split between multiple modes of information through the same channel (e.g., graphics and text, lyrics, speech, etc.), [21]. Concurrently, when information is delivered through separate modalities, such as graphics and speech, retention rates outperformed those with split attention. For example, reading can be considered the learned action of converting visual stimuli (letters) to auditory perception (speech). By delivering the information in speech compared to text, visual attention can focus solely on graphics, compared to graphics and text, while auditory attention can focus on the perception of speech. This type of reception minimizes split attention by decreasing the number of different types of stimuli being encoded through a singular channel, in turn, decreasing the germane load needed to make meaning, [11].

2.2 Distractors in VR

Recent VR research suggests that task-irrelevant stimuli can degrade performance. When irrelevant stimuli are more salient, they are more likely to induce distraction, and impact performance more when a task is easy to perform, [15], [22], [23]. In previous research in display screen contexts, background motion only caused a decrease in task performance during a first-person shooter task compared to a visual searching activity task, [24]. Simultaneously, increased background complexity caused a decrease in task performance regardless of task type.

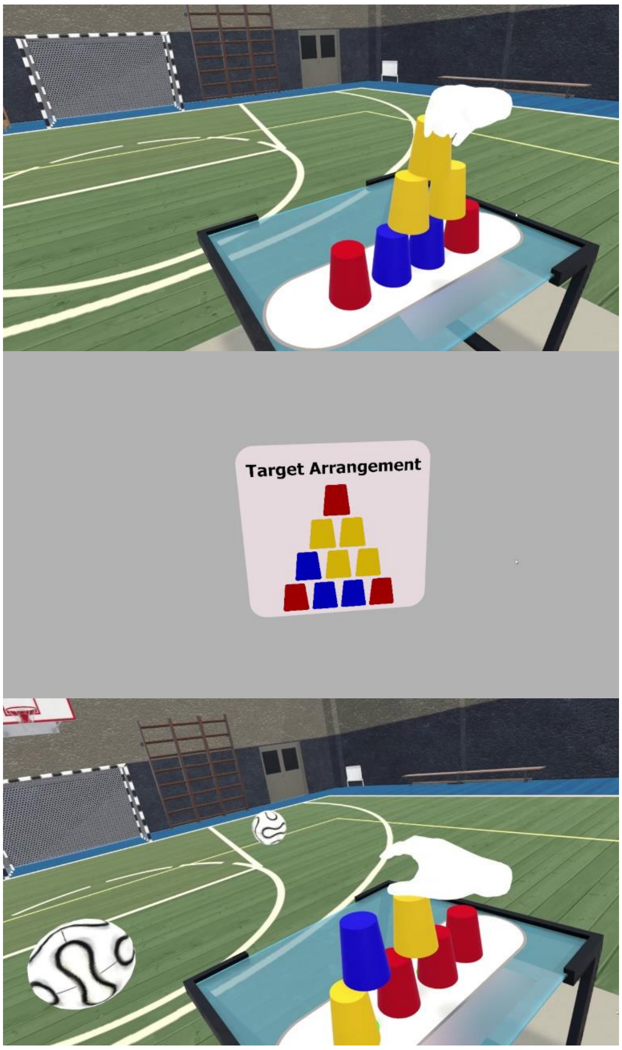

Fig. 1: In-VR screenshots depicting the virtual environment and the task. Top: An in-progress view of the cup-stacking VR task in the No Distractor condition. Middle: The color pattern for this task instance's cup stacking, covering the user's whole view. The user is pressing the side grab button for the pattern to be visible. Bottom: The user is performing the cup-stacking task in the Visual Near distractor condition. Two balls are passing by across the user's FoV.

In HMD contexts, increasing fidelity and other aspects of environmental motion and clutter was suggested to increase the affective quality of the experience yet could overload the visual channel, decreasing task performance. It was suggested that finding a balance for increasing presence, but not cognitive load, would afford the opportunity for flow, increase enjoyment, and increase task performance, [23].

To understand the effects of fidelity, complexity, and task scope on task performance a variety of visual-spatial tasks have been tested in [25]. It was found that increasing environment fidelity significantly increased task performance. Opposingly, when there was an increase in visual complexity and task scope, speed was adversely affected although task performance remained consistent. This suggests that visual clutter, not fidelity, affects speed and task performance in visual-spatial tasks. When clutter and motion are increased in VR and augmented reality (AR) contexts, task performance has been found to decrease, [12], [13]. Medium amounts of background clutter yielded the largest decrease in task performance (i.e., correct answers vs. false positives) when measuring the perceived velocity of objects in VR, whereas low and high amounts of background clutter yielded similar, yet not as severe, results for decreased task performance. Higher visual complexity has also been found to decrease performance when using VR for target detection training, while increasing the field of view (FoV) has been found to increase performance in the same context, [26]. Increasing the FoV would allow for a decrease in the number of fixations that require head movements to make meaning of visual information from different spatial locations (e.g., split attention), lightening the cognitive load. Manipulating complexity, motion, and clutter may decrease task performance when in the same FoV because they would require more fixations to draw meaning from greater visual detail; thus, increasing cognitive load.

The coherence and apprehension principles suggest that all irrelevant graphics, sound, text, and symbols should be removed from the multimedia learning experience to lower the extraneous load on working memory, [14], [16], [27]. However, they have been known to contribute to the affective quality of virtual experiences, [23]. Hence, more research is needed on irrelevant stimuli in VR to understand how various distractions with varying properties affect cognitive load and alertness. In this research, we aim to address this gap by exploring visual and auditory distractors with different levels of intensity and proximity in regard to perceived cognitive load and cognitive alertness within a visuo-spatial task context in VR.

3 The VR Task

3.1 Task Design

It is suggested in the literature that if there are stimuli that are not relevant to the task at hand, it can cause reduced performance. The more salient the stimuli are, the more likely they cause distraction in users. Moreover, when a task is easy, task-irrelevant stimuli are expected to affect user performance more, [15], [22], [23]; and inattentional blindness, which could cause users to ignore distractions, can be typically avoided better, [28], [29]. Based on these, we wanted to design a task that required visual, spatial, and cognitive processing. We designed a cup-stacking VR task that leveraged working memory due to the cognitive-motor nature and the non-difficulty of the cup-stacking task itself. Participants were required to physically manipulate the cups, which would activate the spatial system; visually organize the cups based on their colors, which would activate the visual system; and hold the pattern in their working memory. The same task was used in a previous study of ours with no distractors, [30].

In our study, the participant's objective was to stack 12 virtual cups (4 blue, 4 yellow, and 4 red) in a given color-based pattern (Figure 1). The pattern was randomized for each task instance. For interaction, direct manipulation was used. They could reach a virtual cup by moving the hand-held controller (Figure 2). To grab a virtual cup, they pressed the trigger on the controller. When grabbed, the cup was anchored to the hand in terms of movement and rotation. To release the cup, they released the trigger button. The user could use both controllers to manipulate the cups. The color pattern can be viewed by pressing the side grip button on the controller. When viewed, the pattern overlay covered the whole view of the user. The pattern could be viewed on-demand (was not constantly present). Viewing the pattern paused the task, to require the participants to hold the pattern in their working memory. To submit a pattern configuration, the participant needed to reach and press a virtual button on the table. On average, a task trial (i.e., instance) took 42 seconds (SD = 13.611). When they completed all instances of a trial condition, participants were requested to respond to a few cognitive load-measuring questions on an in-VR questionnaire, before moving on to the next trial condition.

Visual and auditory distractors were introduced in the distractor conditions. The visual distractors included bouncing balls with far, close, and alternating proximity, and the auditory distractors included pure-tone click train sounds with low, high, and alternating pitch. The distractors appeared at the 5 to 10 seconds mark after a task instance started (it was picked randomly at each task instance) and only after the participant viewed the pattern once.

The visual distractors moved horizontally or vertically across the user's FoV. The texture of the ball was randomly picked among three textures: a soccer ball texture, a basketball texture, and a volleyball texture. The pre-assigned travel speed of the balls was increased over time within each task instance. The initial distractors were assigned a horizontal speed of 1 m/s, then the assigned speed was increased linearly to 2 m/s over 60 seconds. After that, each visual distractor in that instance was assigned a horizontal speed of 2 m/s. The ball bounced while traveling from 0 to 1.5 meters (linearly decreased to 1.2 meters over 8 seconds to represent the lost energy). The ball was rotated in its negative z-axis while it traveled. The visual distractors began being triggered within a period of 24 seconds. The period reduced linearly to 12 at the 60-second mark (and the remaining 12 after that if the task instance took longer). The number of traveling balls that simultaneously appeared could be 1, 2, or 3, decided randomly at run time for each triggered distractor within a task instance.

For the audio distractors, pure tone click train sounds that included one-second audio and one-second silence (in a beeping pattern) were used. The low-pitch audio distractors were at 500 Hz, and the high-pitch audio distractors were at 4000 Hz. The audio distractors repeated 1, 2, or 3 times each time they were triggered, decided randomly at run time for each distractor within a task instance. The period of the audio distractors started at 12 seconds and was reduced linearly to 6 over 60 seconds. After that, each audio distractor was assigned a period of 6 seconds. The volume (amplitude) of the audio distractors was increased linearly over 60 seconds until twice the initial value. The VR headset's volume was set at 40% for all participants.

The distractor properties were adjusted through in-house testing and pilot studies with the goal of inducing distractors that would be noticeable but not overwhelming. The visual distractors were triggered less frequently as compared to the audio distractors, since when triggered, the visual distractors remained for a longer duration in the scene while traveling across the FoV of the user.

The virtual environment was designed as a virtual gym for two reasons: (1) to leverage the familiarity of the participants from real-life experiences; (2) to be in harmony with the selected distractors (i.e., balls which can be often seen in gyms and pure tone audio which can be produced by the scoreboard systems in gyms).

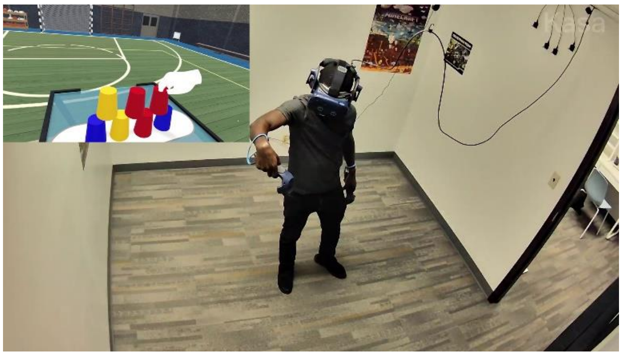

Fig. 2: A user is seen performing the VR task. They are arranging the cups in the given pattern by manipulating the cups. The user is wearing the HTC Vive Pro headset. The headset's cable is fed through a ceiling reel. The user's view from inside the VR headset was added as an overlay on the upper left of the image

3.2 Hardware and Software

We used an HTC VIVE Pro VR headset with eye-tracking [31]. The VR headset was plugged into a desktop computer with an Intel Core i7-6850K processor, NVIDIA GeForce GTX 1080 Ti with 11GB GDDR5X graphics card, 256GB SSD + 2TB 7200RPM hard drives, and 32GB Quad Channel DDR4 memory. The headset cable was fed through a ceiling-mounted reel to prevent tripping over the cable (Figure 2). The sessions were recorded with a video camera and an in-game virtual camera. The software was implemented in Unity [32] using C#. The software ran over 90 frames per second.

4 Evaluation

4.1 Research Questions and Hypotheses

The literature suggests that the perception of cognitive load would increase with the presence of a distractor or with an increase in distractor intensity [1], [11]. Cognitive alertness is also expected to increase with the presentation of stimuli, [33]. Based on these, we constructed the following research questions and hypotheses.

Research Question 1: *"How do distractions affect perceived cognitive load and cognitive alertness in a cognitive-motor dual-task in VR?"*

- Hypothesis 1 (H1): Perceived cognitive load will increase in the presence of a distraction compared to without any distraction.
- Hypothesis 2 (H2): Cognitive alertness will increase in the presence of a distraction compared to without any distraction.
- Hypothesis 3 (H3): Perceived cognitive load will increase more with visual distractions than auditory distractions.
- Hypothesis 4 (H4): Cognitive alertness will increase more with visual distractions than auditory distractions.

Research Question 2: *"How do distraction features affect perceived cognitive load in a cognitive-motor dual-task in VR?"*

- Hypothesis 5 (H5): If the pitch of the auditory distraction is higher, the perceived cognitive load will increase more as compared to the lower pitches.
- Hypothesis 6 (H6): If the visual distraction is closer to the participant, the cognitive load will increase as compared to farther movements.

4.2 Participants

We conducted an IRB-approved within-subjects user study with 50 participants. 2 participants' data were excluded due to not being within the inclusion criteria ($N = 48$). The participants' ages ranged from 18 to 35, $M = 23.12$, $SD = 4.523$. The self-reported gender identification was 46% female, 44% male, 8% non-binary, and 2% other. The participants had little (i.e., less than 1 hour) to no prior VR experience to avoid any pre-established preferences. The participants were undergraduate and graduate students of various majors.

4.3 Procedure

Participants were greeted and offered an informed consent form when they arrived at the laboratory. They were asked to thoroughly read through and ask any questions before signing. Participants completed a brief demographic questionnaire. Then they completed a pre-experience questionnaire including motion sickness questions (out of the scope of this paper). They were escorted to the VR room. The researcher explained the directions of the task verbally. They let the participants know they could stop at any time for any reason if they wished and asked them to complete the tasks as quickly and as accurately as possible (for consistency across participants). Then they were fitted with the VR HMD. To gather a baseline for resting cognitive alertness, the participant was asked to "relax and do nothing" for 7 seconds in an empty virtual room. Once the baseline pupil dilation data was established, the participants were asked to follow the directions in the virtual environment. The participants completed a tutorial with 3 trials first, then the no distractor condition with 6 task trials, followed by the 6 distractor conditions (with 6 trials each) in a randomized order. After each trial condition, an in-VR questionnaire was prompted, and then the resting scene was triggered to collect baseline data for the upcoming trial condition. When all conditions were completed, the researcher helped the participant take off the equipment. Participants then filled out a post-experience questionnaire. The participants were offered a gift card for their participation, and the session ended. The overall VR exposure (from the start of the experiment to the finish) took around 40 minutes on average (*SD* = 8.914).

4.4 Metrics

The participants completed pre and post-questionnaires on a tablet computer. Inside the VR experience, participants were asked questions regarding cognitive load after completing the tutorial and each trial condition. These questionnaires were chosen based on their common use in previous research to analyze cognitive load, [11]. The first set of questions were regarding intrinsic cognitive load and were prompted after the tutorial (learning phase), which was before the experimental condition phases started. These questions were taken from [27] to gather a baseline for the difficulty of a specific task in reference to intrinsic load. Similar questions were asked in between each trial condition to understand the effects each distraction feature had on extraneous cognitive load, [11], [34]. These questions were taken from the NASA TLX, although the full evaluation tool was not employed and only the relevant questions were selected, [35]. The scales used for this research depicted mental demand, temporal demand, effort, and frustration.

Cognitive alertness can be measured with pupil dilation as pupil dilation and cognitive processes have been known to be linked, [36]. Many previous studies also showed a link between pupil dilation and intensity of attention, [37], [38], [39].

Data collected in the software represented the time for each trial's completion, the amount of time gazing at different parts of the environment (i.e., the easel showing the task instructions, the table where the participant performed the task, and the visual distractors passing by), the number of times the participant viewed the pattern, the distractors triggered, and the correctness of the submitted cup pattern. Data from the tutorial instances were discarded from the analyses.

5 Results

5.1 Self-Reported Intrinsic Cognitive Load Results

Perceived intrinsic load was measured through three questionnaire questions asked only once after the tutorial session. The reason for measuring this was to ensure that the task itself was not difficult to execute for the participants in terms of mental or physical demand. The following questions were asked: *"How mentally demanding was the task you just completed?"*, *"How physically demanding was the task you just completed?", "How hard did you have to work to accomplish your level of performance?"*, (answered on a 7-point Likert scale, 1: Very low, 7: Very high). An aggregate percentage score was calculated by adding all response ratings, dividing by the overall possible value (21), and multiplying by 100. The scores can be seen in Table 1. The overall perceived intrinsic load was on the lower end for the aggregate score out of 100% (*M* = 29.07, *SD* = 11.25), and was on the lower end of the 7-point Likert scale for all three measures (i.e., mental demand, physical demand, and effort). This indicates that the task was not difficult to execute, as it was intended in our study.

Table 1. Average Intrinsic Cognitive Load Scores

	Aggregate Score %	Mental Demand	Physical Demand	Effort
Mean	29.07	2.21	1.65	2.25
SD	11.25	1.17	0.86	1.00

5.2 Self-Reported Extraneous Cognitive Load Results

To measure perceived extraneous cognitive load, we asked the following questions after completion of each of the seven trial conditions in VR: *"How mentally demanding was the task you just completed?"*, *"How hard did you have to work to accomplish your level of performance?"*, *"How stressed, discouraged, irritated, and annoyed were you?"*, *"How hurried or rushed did you feel during this task?"*. Participants answered on a 7-point Likert scale, 1: Very low, 7: Very high. The scores were calculated in the same fashion as intrinsic cognitive load. Average scores for all conditions can be seen in Table 2. The following abbreviations will be used for the seven trial conditions: ND: No Distractor, AA: Audio Alternate, AH: Audio High, AL: Audio Low, VA: Visual Alternate, VC: Visual Close, VF: Visual Far.

A repeated measures ANOVA test was run on the aggregate scores of all seven trial conditions,

$F(4.108, 193.096) = 13.472$, $p <$ **0.001**, indicating a statistical significance. Sphericity was not assumed in Mauchly's test and Greenhouse-Geisser correction was used. When paired sample T-tests were performed between the No Distractor condition and each of the distractor conditions individually, there was a statistically significant difference between each audio distractor condition vs. the No Distractor condition. There was not any statistical significance in the visual conditions. We also compared the overall average of the visual distractor conditions with the overall average of the audio distractor conditions to investigate H3; Audio High vs. Audio Low to investigate H5; and Visual Close vs. Visual Far to investigate H6 (Table 3 for detailed T-test results).

Table 2. Average Extraneous Cognitive Load Scores for All Conditions (Aggregate Score %)

	ND	AA	AH	AL	VA	VC	VF
Mean	43.38	54.99	54.99	50.00	43.38	43.30	42.49
SD	13.80	18.74	19.98	18.00	17.77	17.19	15.81

Table 3. Paired Samples T-Test Results for Scaled Extraneous Cognitive Load Scores

Condition	*t*	*df*	*p*	Mean Difference
ND v. AA	-5.087	47	< **0.001**	-11.607
ND v. AH	-4.433	47	< **0.001**	-10.863
ND v. AL	-2.993	47	**0.004**	-6.622
ND v. VA	0.000	47	1.000	0.000
ND v. VC	0.033	47	0.974	0.074
ND v. VF	0.435	47	0.666	0.893
AH v. AL	2.364	47	**0.022**	4.241
VC v. VF	0.504	47	0.617	0.818
Audio v. Visual	4.542	47	< **0.001**	7.813

5.3 Self-Reported Level of Distraction Results

To understand the participants' perceived level of distraction, in the post-interaction questionnaire, which was filled out when the participant completed all seven trials, we asked the following questions:

1. *"Were you distracted by the ball?"*
2. *"Rate your level of distraction caused by the ball."*
3. *"Were you distracted by the texture of the ball changing?"*
4. *"Were you more distracted when the ball was closer or farther away?"*
5. *"Do you think there was a change in the amount of speed the ball was moving within a level?"*
6. *"Do you think there was a change in the direction the ball was moving?"*
7. *"Were you more distracted as the ball moved horizontally or vertically?"*
8. *"Were you distracted by the sound?"*
9. *"Rate your level of distraction caused by the sound."*
10. *"Which level of pitch was more distracting?"*
11. *"Do you think there was a change in the level of amplitude of the sound within a level?"*
12. *"Do you think there was a change in how frequently the sound was played within a level?"*

We added questions 4 and 10 after having 12 participants (based on observations), hence the remaining 38 participants' data was used for analyzing the responses to these questions.

5.3.1 Self-Reported Visual Distraction

50% of the participants stated that they were distracted by the visual distractor (i.e., balls). When asked to rate their level of distraction from the ball (1: Very Low, 7: Very High), the average response was 3.13, $SD = 1.329$. Some participants suggested they only noted the ball until they were aware it would not interfere with the task. Some participants mentioned being more distracted at the onset of the visual distractor entering the environment. Among the participants who stated being distracted by the ball, 79% suggested they were most distracted by the ball when it was closer compared to farther away.

92% of the participants stated that they were not distracted by the changed textures on the ball. Some participants suggested that they were not distracted by the changes in texture because the ball did not stay in their field of view long enough to note the change in texture from one pass to another.

32% of the participants noticed that there was a change in the speed of the ball's movement. Of those, approximately one-third felt distracted by it. One participant noted, *"The speed of the ball seemed to increase which naturally increases the seeming chaos of the space."* Another suggested, *"That is when I was worried that they could hit my table."* Also, 68% of the participants suggested they were distracted by the changes in the directionality of the balls. Of those participants, 50% noted they were more distracted as the ball was moving vertically compared to 50% suggesting horizontal movement more distracting.

5.3.2 Self-Reported Auditory Distraction

77% of participants were distracted by the audio distractors with an average distraction rating of 5.50 out of 7, *SD* = 1.425. Participants suggested the sound interrupted their concentration and train of thought. They mentioned, *"I was saying the colors out loud so that I could remember better but the sound was preventing me from doing that."* or *"It reminded me of the seatbelt chime that some cars have and I just wanted it to stop."* 86% of the participants who found the audio distracting suggested that the high-pitch frequencies were more distracting than the low-pitch frequencies. One participant suggested, *"Maybe because higher pitches are used for alarms so they're more distracting and set off some flight or fight reaction."*

We compared the self-reported visual distraction vs. auditory distraction based on the responses to questions 1 and 8 (answered yes or no; 1 indicating yes and 0 indicating no). We conducted a Wilcoxon signed ranks test, and there was a statistical significance ($Z = -3.153$, $p =$ **0.002**). Participants reported being distracted by the audio distractors more ($M = 0.77$, $SD = 0.425$) as compared to the visual distractors ($M = 0.50$, $SD = 0.505$).

57% of the participants noticed a change in amplitude with the audio distractors and 76% of those participants suggested they were more distracted as the sounds got louder and were in higher pitch. Participants mentioned, *"As it got louder, it became more difficult to concentrate and I had to check my work more frequently."* or *"If it stayed the same, I would have become more used to it, but when it changed, it grabbed my attention all over again."* 75% of participants noted there was an increase in the frequency of the audio distractors over time. Of those participants, 75% suggested they were distracted by it. They noted, *"I was more distracted when the sounds played more frequently because it grew increasingly annoying and jarring hearing the same sound over and over again."* or *"They stopped my train of thought more often than any other change in the environment. It engaged too many of my senses, and I would have to rebuild those thoughts more frequently because of it."*

5.4 Eye-Tracking Data Results

5.4.1 Pupil Dilation

Cognitive alertness can be detected by event-related changes that follow an alerting signal to the cognitive system, [40]. This is different than cognitive load, which relates to the amount of working memory resources used. In [40], it was found that pupil dilation occurred in tandem with cognitive alertness. Therefore, to investigate cognitive alertness, we measured pupil dilation (PD) event-related changes after environmental stimuli. The pupil dilation data was collected with the HTC VIVE Pro Eye VR headset. To understand the event-related changes in pupil dilation, the following equation was used in our study for the averages of both eyes' pupil dilations: *PD Change = (Test interval pupil dilation - Baseline interval pupil dilation) / Baseline interval pupil dilation.* For the baseline interval, baseline pupil dilation data was collected at the beginning of each trial condition in the resting scene in an empty virtual room for each individual. For the test interval, the pupil dilation data between 0.1 and 2 seconds after the initiation of a distractor was used. These values were chosen in alignment with previous research as pupil dilation has been shown to react to changes in cognitive load relatively quickly, [41].

The average PD change values for both eyes can be seen in Table 4. No statistical significance was found in the repeated-measures ANOVA tests of the overall average pupil dilation ($F(6, 282) = 0.909$, $p = 0.489$). Sphericity was assumed in Mauchly's test.

We conducted pre-planned comparisons to investigate our relevant hypothesis and compared the overall average of visual distractors with the overall average of audio distractors with a Paired samples T-test to investigate H4. There was no statistical significance ($t(47) = 0.408$, $p = 0.685$).

Table 4. Average Pupil Dilation Change Values for Both Eyes for Different Trial Conditions

	ND	AA	AH	AL	VA	VC	VF
Mean	-0.0204	0.0037	-0.0053	-0.0079	-0.0038	-0.0002	-0.0179
SD	0.0781	0.1051	0.1015	0.1042	0.1069	0.1055	0.1012

5.4.2 Gaze Duration

Data was collected through our software for how long each participant gazed at different areas of relevance during the task: the ball (distractor), easel (written instructions), and the table (task area).

Table 5. Average Gaze Duration for Different Areas of Interest and Total Duration (in seconds)

	ND	AA	AH	AL	VA	VC	VF
Ball	-	-	-	-	2.05	3.11	0.85
Easel	29.26	15.80	13.91	13.31	16.67	15.14	14.91
Table	171.48	139.02	136.47	133.62	128.82	131.24	131.82
Total	308.60	245.54	244.00	242.58	240.56	239.94	244.04

Table 6. Average Normalized Gaze Duration (in percentage)

	ND	AA	AH	AL	VA	VC	VF
Ball	-	-	-	-	0.87	1.33	0.26
Easel	9.62	5.82	5.81	5.52	6.82	7.02	6.03
Table	54.88	56.55	55.95	55.97	54.18	56.79	55.00

These areas can be seen in Figure 3. The sum of the amount of time gazed at each location was taken for each instance and averaged for a score that represented each condition (Table 5).

Since instance duration differed for trial conditions, we normalized this data by dividing it by the total duration of instances in each trial condition (excluding the tutorial instances). The normalized durations can be seen in Table 6. Note that the total might not be 100% since the participants could also look elsewhere in the virtual environment. On average, the participants gazed at the ball longer in the Visual Close condition (3.11 sec, 1.33% of the total time), followed by the Visual Alternate condition (2.05 sec, 0.87% of the total time), and the Visual Far condition (0.85 sec, 0.26% of the total time). A statistical significance was found when a repeated measures ANOVA test was run on the normalized gaze duration (percentage) for the ball distractor ($F(2, 94) = 22.801$, $p < \mathbf{0.001}$). Sphericity was assumed in Mauchly's test. When paired-sample T-tests were performed on the data, a statistically significant difference was found for all three comparisons: Visual Alternate and Visual Close ($t(47) = -2.766$, $p = \mathbf{0.008}$), Visual Alternate and Visual Far ($t(47) = 4.111$, $p < \mathbf{0.001}$), and Visual Close and Visual Far ($t(47) = 6.626$, $p < \mathbf{0.001}$). See Table 7 for detailed results.

Repeated measures ANOVA tests were also run on the normalized gaze duration (percentage) for the Easel (instructions) and the table (task area). Statistical significance was found for Easel gaze duration: $F(2.422, 113.840) = 3.840$, $p = \mathbf{0.018}$. There was no statistical significance for table gaze duration: $F(1.476, 69.389) = 0.249$, $p = 0.711$). Sphericity was not assumed in Mauchly's test for both Easel and table data and Greenhouse-Geisser correction was used (Table 7 for details on the paired-sample T-test statistics for Easel gaze duration).

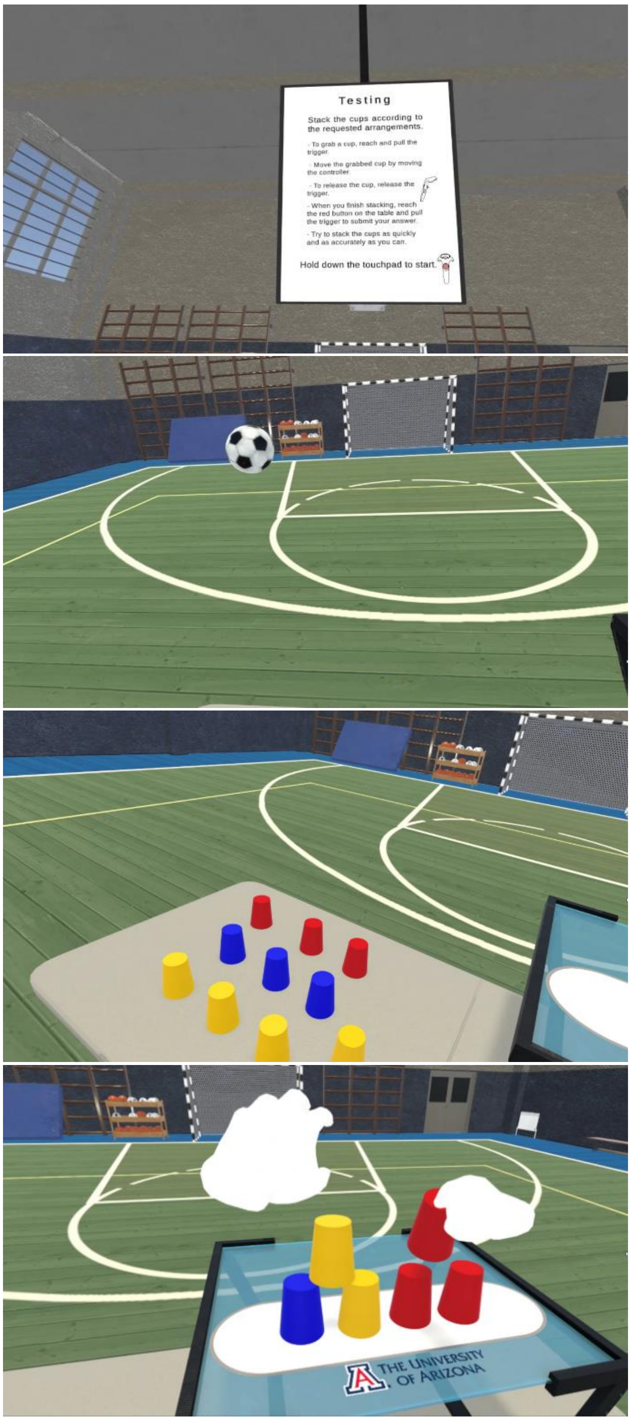

Fig. 3: In-VR screenshots depicting the target areas analysed for Gaze Duration. In the first image from the top, Easel with the instructions can be seen. The second image shows the user looking at the ball. The third and fourth images show the task area spread across the table. The third image shows the cups at the beginning of each instance and the fourth image shows the cups being stacked mid-instance. In both the 3rd and 4th images, the user is looking at the table

Table 7. Paired Samples T-test Results for Normalized Gaze Duration (Percentage) at Relevant Fixation Locations (Ball and Easel)

Area	**Condition**	***t***	***df***	***p***	**Mean Diff.**
Ball	VA v. VC	-2.766	47	**0.008**	-0.461
	VA v. VF	4.111	47	< **0.001**	0.609
	VC v. VF	6.626	47	< **0.001**	1.070
Easel	ND v. AA	5.186	47	**< 0.001**	3.806
	ND v. AH	4.754	47	**< 0.001**	3.814
	ND v. AL	5.912	47	**< 0.001**	4.107
	ND v. VA	2.893	47	**0.006**	2.806
	ND v. VC	1.581	47	0.121	2.605
	ND v. VF	4.666	47	**< 0.001**	3.596

6 Discussion

6.1 Perceived Cognitive Load with Distractors (H1)

The literature suggests that both reception and perception of cognitive load would increase with the presence of a distractor or with an increase in distractor intensity, [1], [11]. We predicted that the perception of extraneous cognitive load, surveyed through the in-VR questionnaires filled out after completing each trial condition, would increase in the presence of distractions compared to performing the same task in the same virtual environment without any distractions. We found the perceived extraneous cognitive load was higher in the audio distractor conditions than in the No Distractor condition. However, there was no statistically significant difference between the visual distractor conditions and the No Distractor condition. The post-interaction survey results suggest that H1 is supported for the audio distractor conditions but is not supported for the visual distractor conditions.

These subjective evaluations were also in alignment. 77% of the participants were distracted by the audio distractors whereas only 50% were distracted by the visual distractors. When asked to rate their level of distraction, the rating was 5.5 out of 7 for the audio and 3.13 for the visual distractor (7: very distracting). The perceived level of distraction from the visual distractors was low, but participants perceived to be distracted by the audio distractors, finding it a hindrance to their task performance. Designers should be careful when adding audio distractions similar to beeping sounds in VR as this may increase the perceived cognitive load in users, especially with higher pitches.

6.2 Cognitive Alertness with Distractors (H2)

In [40], it was found that cognitive alertness, measured through pupil dilation, correlates with cognitive load. Similarly, we predicted that cognitive alertness would increase in the presence of a distraction. There was no statistically significant difference in the ANOVA comparing the pupil dilation data for all seven trial conditions. Hence, H2 was not supported. Although confirmatory studies are needed, a possible reason could be the difficulty level of the task. We intentionally picked our task to be moderately challenging in terms of difficulty level. Thus, participants might not have been affected by the audio distractors in terms of cognitive alertness due to the cognitive load increase not being above a threshold to be detected with biological markers. There is a mismatch between the perception of cognitive load and cognitive alertness in our study, which should be investigated more in future studies using subjective and biological measurements.

6.3 Split Attention with Visual Distractors (H3 and H4)

In [14], it was found that splitting visual attention between two locations increases perceived cognitive load. Since the task we used required the use of the visual channel for completion, we predicted that the visual distractions would also split visual attention. Moreover, previous research regarding perceived cognitive load suggests that the split attention caused by the visual distractors would lead to a greater increase in perceived cognitive load, especially if movement and speed are increased across time, [12], [21].

Two forms of perception data were collected to analyze if the extraneous cognitive load would increase more with visual distractors than with auditory distractors. We took an average of the intermittent in-VR extraneous cognitive load question results for the visual conditions, and the audio conditions, and compared the two with a paired samples T-test. The Audio Alternate condition increased perceived extraneous cognitive load from the No Distractor condition by 14%, followed closely by the Audio High condition with a 13% increase, and the Audio Low condition with a 9% increase. The visual conditions ranged from a 1-2% increase in perceived cognitive load notably lower than the perceived load in the audio distractor conditions.

We also compared the results for the questions: *"Rate your level of distraction caused by the ball."*

vs. *"Rate your level of distraction caused by the sound."* (1: Very Low, 7: Very High) asked in the post-interaction questionnaire. This self-reported perception data suggested the participants perceived the audio conditions as more distracting (78%) than the visual conditions (48%).

In the post-interaction questionnaire's open-ended feedback section, 28% of participants suggested they "noticed" the ball until they were "aware" that it would not interfere with the task at hand. Once the ball was perceived as irrelevant to the task, the participants did not feel it affecting their cognitive load. Some participants suggested they were distracted by the sound due to its resemblance to an alarm sound. These results could be due to the affordance of sound to draw focus with its many uses in our everyday lives as signals to get our attention. Moreover, the audio tones used in our study were pure tones and resembled an emergency or safety signal that is often used in sirens, house alarms, smoke detectors, etc. Since we as humans have trained our cognition to perceive these sounds readily in our past experiences, the audio conditions in our study could have been more salient to the participants than the visual conditions. Our results do not support H3.

Cognitive alertness data was simultaneously analyzed. We took an average of the change in pupil dilation for the visual distractor conditions and the audio distractor conditions and compared the two with a paired samples T-test. There was no statistically significant difference. Therefore, our H4 was also rejected.

6.4 Pitch (H5)

We hypothesized that the perceived cognitive load would increase more with the higher pitches as compared to the lower pitches (H5). We analyzed this based on the in-VR questionnaire and the post-interaction questionnaire results. The paired samples T-test comparing the Audio High and Audio Low distractor conditions resulted in statistical significance. Participants perceived their cognitive load to be higher with the Audio High distractor condition. According to the post-interaction questionnaire, 86% of participants perceived the higher-pitch tones to be more distracting than the lower-pitch tones. These findings are in alignment with previous research [42] suggesting that higher pitches increase perceived cognitive load, supporting our H5.

One participant suggested that the higher pitch tones drew attention more readily as follows: *"Higher pitches are used for alarms so they're more distracting and set off some flight or fight reaction."* These results could be allocated to the use of these pitches in alarms and our conditioned responses to these signals. More specifically, higher pitches are often used to signal urgency (e.g., sirens, fire alarms, etc.). Based on these findings, designers should only use high-pitch sounds when they actively trying to signal the user to make an action or stop an action they are attempting to complete. The use of high-pitch tones in other contexts may lead to unwanted confusion because of the increases in perceived cognitive load and may potentially decrease the enjoyment of the VR experience.

6.5 Proximity of Visual Distractors (H6)

The perception of proximity of the visual distractors was analyzed through the perceived extraneous cognitive load in the in-VR questions asked and the post-interaction questionnaire data. We compared the data for Visual Close vs. Visual Far conditions with a paired-sample T-test for the perceived extraneous cognitive load. The score was slightly higher in the Visual Close condition than in the Visual Far condition, although there was no statistical significance. Thus, H6 is not supported. The reason could be the visual distractors were task-irrelevant. In the open-ended questionnaire responses, many participants noted being worried the ball would interfere with the task or the table they were performing the task on (i.e., cup stacking) and when they learned it would not, they paid little of their perceived attention. Including task-relevant visual distractors may change perceived attention for future studies.

We also examined the post-interaction questionnaire data. We considered the responses to the question *"Were you more distracted when the ball was closer or farther away?"*. Among the ones who were distracted by the visual distractors, 79% suggested they were more distracted by the ball when it was closer, similar to previous research, [43]. However, only 50% of the participants stated that they were distracted by the visual distractor in our study. Consequently, these results are not strong enough to support our H6.

Although the perceived cognitive load was not affected significantly, the proximity of visual distractions affected the gaze duration statistically significantly in our study. The participants gazed at the ball for at least 1 second longer in the Visual Close condition compared to the Visual Far or Visual Alternate conditions. This indicates that visual distractions attracted more attention when they were closer to the participants. Future developers should be aware that increasing the proximity of visual distractions can attract the attention of the users more and may cause an increase in task completion duration.

6.6 Speed and Frequency

During each trial instance, the intensity of speed and amplitude increased across time, although pitch and proximity remained consistent. This was to promote the participants being more distracted over time due to the increase in distractor intensity. We analyzed the post-interaction questionnaire data, using the responses to the following question: *"Were you more distracted when the speed of the ball increased? Please explain."* The results suggested that the speed of the visual distractions was perceived by ~30% of the participants and only less than 10% felt distracted by it.

Participants' perceived level of distraction was also not affected by the changes in the texture of the ball (only 8% mentioned being distracted by the texture change). On the other hand, most of the participants suggested that they were distracted by the changes in the directionality of the balls. No notable effect of the specific direction of the movement (i.e., horizontal or vertical) was observed, although the change itself in directionality seemed to cause distraction in the participants.

Participants also noticed the increase in the frequency of the audio distractions more, where 75% of the participants noticed the increase in the audio frequencies, and ~50% were distracted by it. Several participants noted that the increase in the audio frequencies disrupted their train of thought, especially as it became more frequent. They also suggested feeling more hurried and rushed as the sounds became more frequent, which was in alignment with previous findings, [15]. This increase in the sense of urgency in the sound distractor conditions is likely due to the use of audio tones to signal actions in everyday interaction. We are conditioned to complete physical actions when we hear sounds in our lives (e.g., seat belt chime, bell at a reception office, etc.). More importantly, we are classically conditioned to feel a sense of urgency until we make the sound go away by completing the action, [44]. Increasing the frequency of the audio click trains could have created a greater sense of urgency and the need to complete the action quickly. In the open-ended responses, many of the participants also mentioned disliking the sound in general, furthering the want to make it stop by completing an action.

In summary, the results suggested that participants were not affected by the increases in speed in the visual distractions across time, yet they were perceiving the increases in the click train frequencies in the audio distractor conditions that can implicitly increase cognitive load. A possible reason behind the difference in the observed perception results in the visual distractions could be the lack of clutter in our research. In this research, we kept the virtual environment non-cluttered and included a few visual distractions to measure the effects precisely. This may not have been enough to distract the participants in a similar way to the previous research where more cluttered environments with distractions were explored, [12], [13]. Our results suggest that the intensity of speed is not perceived as salient in the visual channel compared to the auditory channel when it is incrementally increased across time.

6.7 Amplitude of Auditory Distractors

The amplitude of the audio distractors was increased within each task instance. The post-interaction questionnaire results suggested that over half the participants (57%) noted an increase in the amplitude of the auditory distractors and two-thirds of them (76%) were distracted by it. Several participants also mentioned in the post-interaction questionnaire that the sound broke their concentration, especially as it got louder. These findings are in alignment with previous research suggesting that an increase in amplitude increases the saliency of an auditory stimulus, [42], [43]. Designers of future VR experiences should be careful with the amplitude of audio cues, as the change in the amplitude may cause an increase in distraction.

7 Limitations and Future Work

Our research is limited by several factors, mainly task dependence and visual distractor dependence. We selected the task as cup-stacking due to its visuospatial nature and moderately challenging level of difficulty. This may limit the generalizability of our results to real-world VR applications. It is important to measure similar metrics during more cognitively demanding or multitasking tasks in VR. For the visual distractors, we used virtual balls in a non-cluttered setting. Other types, forms, and sizes of objects can be explored in the future in varying levels of intensity and in virtual environments with varying levels of clutter. The mismatch between subjective and physiological data in our research warrants future studies investigating pupil dilation as a measure of cognitive alertness. Other notable future research directions include different types of audio distractors and their effects on an audio-spatial task compared to the visual-spatial task used in this study. Specifically, research using real-world tones like cell phone alarms, seat belt chimes, or sounds found in nature and their effects on focus based on an individual's funds of knowledge would be important.

8 Conclusion

In this paper, we explored visual and auditory distractors' effects on perceived cognitive load and cognitive alertness within a visuo-spatial cup stacking task context in VR. We measured subjective data along with eye tracking in a within-subjects user study (N = 48). Participants were distracted by the auditory distractors and not the visual distractors. There was a notable increase in the perceived level of distraction when the frequency of the auditory distractions increased, and when the pitch was high. Also, participants gazed at the visual distractors more when they were closer to them. Future developers should be wary of using the auditory channel for information not related to the task at hand unless the goal is to increase the challenge. These results can help guide future designers and developers on stimuli inclusion in VR for more effective learning and/or training experiences.

References:

[1] P. Antonenko, F. Paas, R. Grabner, and T. Van Gog, "Using electroencephalography to measure cognitive load," *Educational psychology review*, 22, 425-438, 2010. DOI: 10.1007/s10648-010-9130-y.

[2] N. Cowan, "The magical number 4 in short-term memory: A reconsideration of mental storage capacity," *Behavioral and brain sciences*, 24, no. 1, 87-114, 2001. DOI: 10.1017/s0140525x01003922.

[3] S. McCloud, *Understanding comics: The invisible art*, 1st HarperPerennial ed., New York: HarperPerennial, 1994.

[4] H. J. Hsia, "The information processing capacity of modality and channel performance," *AV Communication Review*, 19, no. 1, 51-75, 1971. DOI: 10.1007/BF02768431.

[5] G. Makransky, T. S. Terkildsen, and R. E. Mayer, "Adding immersive virtual reality to a science lab simulation causes more presence but less learning," *Learning and instruction*, 60, 225-236, 2019. DOI: 10.1016/j.learninstruc.2017.12.007.

[6] S. Jang, J. M. Vitale, R. W. Jyung, and J. B. Black, "Direct manipulation is better than passive viewing for learning anatomy in a three-dimensional virtual reality environment," *Computers & Education*, 106, 150-165, 2017. DOI: 10.1016/j.compedu.2016.12.009.

[7] J. Parong and R. E. Mayer, "Cognitive consequences of playing brain-training games in immersive virtual reality," *Applied Cognitive Psychology*, 34(1), 29-38, 2020. DOI: 10.1002/acp.3582.

[8] R. K. Lowe, "Animation and learning: Selective processing of information in dynamic graphics," *Learning and instruction*, 13(2), 157-176, 2003. DOI: 10.1016/S0959-4752(02)00018-X.

[9] T. Burke and B. Rooney, "Multi-modal dual-task measurement: A new virtual reality for assessment," *Frontiers in Psychology*, 11, 635413, 2021. DOI: 10.3389/fpsyg.2020.635413.

[10] P. Albus, A. Vogt, and T. Seufert, "Signaling in virtual reality influences learning outcome and cognitive load," *Computers & Education*, 166, 104154, 2021. DOI: 10.1016/j.compedu.2021.104154.

[11] S. Baceviciute, A. Mottelson, T. Terkildsen, and G. Makransky, "Investigating representation of text and audio in educational VR using learning outcomes and EEG," In *Proceedings of the 2020 CHI conference on human factors in computing systems, Honolulu, HI, USA*, pp. 1-13, 2020. DOI: 10.1145/3313831.3376872.

[12] L. Bozgeyikli, A. Raij, S. Katkoori, and R. Alqasemi, "Effects of Environmental Clutter and Motion on User Performance in Virtual Reality Games," *CEUR Workshop Proceedings, Austin, Texas, USA,* vol. 1715, 2016.

[13] V. Ferrer, Y. Yang, A. Perdomo, and J. Quarles, "Consider your clutter: Perception of virtual object motion in AR," In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Adelaide, SA, Australia,* pp. 1-6, IEEE, 2013. DOI: 10.1109/ISMAR.2013.6671835.

[14] R. E. Mayer, "Research-based principles for designing multimedia instruction," in *Applying the science of learning in education: Infusing psychological science into the curriculum,* V. A. Benassi, C. E. Overson, & C. M. Hakala (Eds.), 59-70, 2014.

[15] R. E. Mayer, “Cognitive theory of multimedia learning,” *The Cambridge handbook of multimedia learning*, 41, 31-48, 2005. DOI: 10.1017/CBO9780511816819.004.

[16] R. E. Mayer and L. Fiorella, “12 principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles,” In *The Cambridge handbook of multimedia learning*, vol. 279, pp. 279-315. New York, NY: Cambridge University Press, 2014. DOI: 10.1017/CBO9781139547369.015.

[17] J. Clark, "Exploring the Effects of Multi-Sensory Extraneous Load on Attention and Task Performance for Training Contexts in Virtual Reality." Doctoral Dissertation, University of Arizona, 2023.

[18] J. Clark and L. Boz, “Towards Using Biometric Transcripts for Analyzing Multimodal Sensory Information Processing Data.” To appear in *Proceedings of the International Conference on Human-Computer Interaction*, *Gothenburg, Sweden,* 2025.

[19] S. Sanei and J. A. Chambers. *EEG signal processing*, John Wiley & Sons, 2013.

[20] P. Sörqvist, A. Nöstl, and N. Halin, “Working memory capacity modulates habituation rate: Evidence from a cross-modal auditory distraction paradigm,” *Psychonomic Bulletin & Review*, 19, 245-250, 2012. DOI: 10.3758/s13423-011-0203-9.

[21] R. Moreno and R. E. Mayer, “Visual presentations in multimedia learning: Conditions that overload visual working memory,” In *Visual Information and Information Systems: Third International Conference*, *VISUAL’99 Amsterdam, The Netherlands*, June 2–4, 1999 Proceedings 3, pp. 798-805, Springer Berlin Heidelberg, 1999. DOI: 10.1007/3-540-48762-X_98.

[22] J. Jerald, *The VR book: Human-centered design for virtual reality*, Morgan & Claypool, 2015. DOI: 10.1145/2792790.

[23] I. Kartiko, M. Kavakli, and K. Cheng, “Learning science in a virtual reality application: The impacts of animated-virtual actors’ visual complexity,” *Computers & Education,* 55, no. 2, 881-891, 2010. DOI: 10.1016/j.compedu.2010.03.019.

[24] L. Caroux, L. Le Bigot, and N. Vibert, “Impact of the motion and visual complexity of the background on players' performance in video game-like displays,” *Ergonomics*, 56, no. 12, 1863-1876, 2013. DOI: 10.1080/00140139.2013.847214.

[25] F. Bacim, E. D. Ragan, S. Scerbo, N. F. Polys, M. Setareh, and B. D. Jones, “The effects of display fidelity, visual complexity, and task scope on spatial understanding of 3D graphs,” In *Graphics Interface*, vol. 2, no. 1, pp. 25-32, 2013. DOI: 10.5555/2532129.2532135.

[26] E. D. Ragan, D. A. Bowman, R. Kopper, C. Stinson, S. Scerbo, and R. P. McMahan, “Effects of field of view and visual complexity on virtual reality training effectiveness for a visual scanning task,” *IEEE transactions on visualization and computer graphics*, 21, no. 7, 794-807, 2015. DOI: 10.1109/TVCG.2015.2403312.

[27] P. Ayres, “Using subjective measures to detect variations of intrinsic cognitive load within problems,” *Learning and instruction*, 16, no. 5, 389-400, 2006. DOI: 10.1016/j.learninstruc.2006.09.001.

[28] N. Lavie, A. Hirst, J. W. De Fockert, and E. Viding, “Load theory of selective attention and cognitive control,” *Journal of experimental psychology: General*, 133, no. 3, 339, 2004. DOI: 10.1037/0096-3445.133.3.339.

[29] K. Bredemeier and D. J. Simons, “Working memory and inattentional blindness,” *Psychonomic Bulletin & Review*, 19, 239-244, 2012. DOI: 10.3758/s13423-011-0204-8.

[30] L. Bozgeyikli, E. Bozgeyikli, C. Schnell, and J. Clark, "Exploring horizontally flipped interaction in virtual reality for improving spatial ability." *IEEE Transactions on Visualization and Computer Graphics*, 29.11, 4514-4524, 2023. DOI: 10.1109/TVCG.2023.3320241.

[31] VIVE Pro Eye Overview, [Online]. https://www.vive.com/sea/product/vive-pro-eye/overview/ (Accessed Date: April 28, 2024).

[32] Unity Real-Time Development Platform, [Online]. https://unity.com/ (Accessed Date: April 28, 2024).

[33] E. H. Hess and J. M. Polt, “Pupil size as related to interest value of visual stimuli,” *Science*, 132(3423), pp.349-350, 1960. DOI: 10.1126/science.132.3423.349.

[34] F. G. Paas, “Training strategies for attaining transfer of problem-solving skill in statistics: a cognitive-load approach,” *Journal of educational psychology*, 84, no. 4, 429, 1992. DOI: 10.1037/0022-0663.84.4.429.

[35] S. G. Hart, "NASA-task load index (NASA-TLX); 20 years later," In *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, no. 9, pp. 904-908, Sage CA: Los Angeles, CA: Sage publications, 2006. DOI: 10.1177/154193120605000909.

[36] E. H. Hess and J. M. Polt, "Pupil size in relation to mental activity during simple problem-solving," *Science*, 143, no. 3611, 1190-1192, 1964. DOI: 10.1126/science.143.3611.1190.

[37] S. Ahern and J. Beatty, "Pupillary responses during information processing vary with Scholastic Aptitude Test scores," *Science*, 205, no. 4412, 1289-1292, 1979. DOI: 10.1126/science.472746.

[38] B. Laeng, M. Ørbo, T. Holmlund, and M. Miozzo, "Pupillary stroop effects," *Cognitive processing*, 12, 13-21, 2011. DOI: 10.1007/s10339-010-0370-z.

[39] T. Piquado, D. Isaacowitz, and A. Wingfield, "Pupillometry as a measure of cognitive effort in younger and older adults," *Psychophysiology*, 47, no. 3, 560-569, 2010. DOI: 10.1111/j.1469-8986.2009.00947.x.

[40] A. Petersen, A. H. Petersen, C. Bundesen, S. Vangkilde, and T. Habekost, "The effect of phasic auditory alerting on visual perception," *Cognition*, 165, 73-81, 2017. DOI: 10.1016/j.cognition.2017.04.004.

[41] R. Montefusco-Siegmund, M. Schwalm, E. R. Jubal, C. Devia, J. I. Egaña, and P. E. Maldonado, "Alpha EEG activity and pupil diameter coupling during inactive wakefulness in humans," *eneuro*, 9, no. 2, 2022. DOI: 10.1523/ENEURO.0060-21.2022.

[42] P. D. Skosnik, G. P. Krishnan, and B. F. O'Donnell, "The effect of selective attention on the gamma-band auditory steady-state response," *Neuroscience letters*, 420, no. 3, 223-228, 2007. DOI: 10.1016/j.neulet.2007.04.072.

[43] C. Healey and J. Enns, "Attention and visual memory in visualization and computer graphics," *IEEE transactions on visualization and computer graphics*, 18(7), 1170-1188, 2011. DOI: 10.1109/TVCG.2011.127.

[44] E. V. Clark, "How language acquisition builds on cognitive development," *Trends in cognitive sciences*, 8(10), 472-478, 2004. DOI: 10.1016/j.tics.2004.08.012.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

- Jack Clark carried out the experiment design, user studies, data collection, analysis, and writing as a PhD student.
- Lila Boz oversaw the research and writing as the PhD advisor.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare.