# The ears point to the eyes

Tiago Teixeira*
tiago.renou@tecnico.ulisboa.pt
Instituto Superior Técnico
Student Nr. 102638
Portugal

João Ribeiro*
joao.r.ribeiro@tecnico.ulisboa.pt
Instituto Superior Técnico
Student Nr. 102715
Portugal

João Coelho*
joaopgcoelho@tecnico.ulisboa.pt
Instituto Superior Técnico
Student Nr. 103141
Portugal

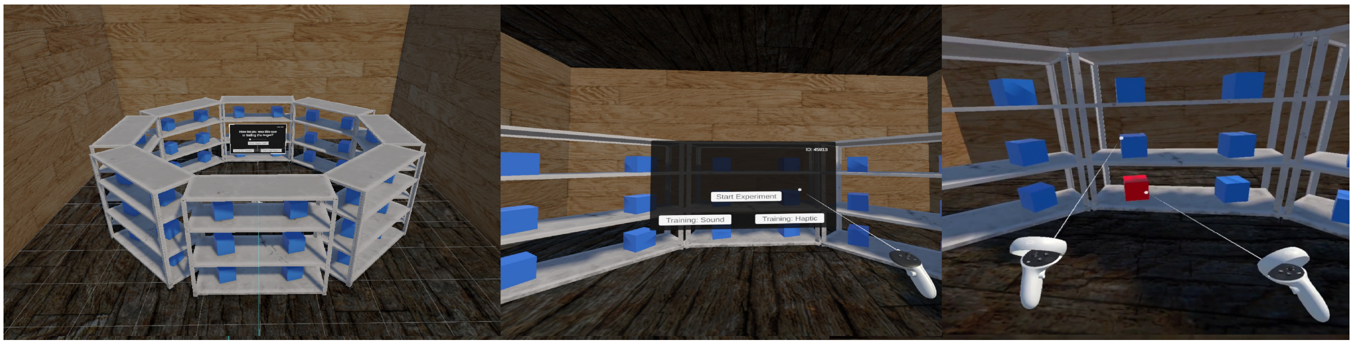



**Figure 1: Virtual environment used for the experiment**

## Abstract

Navigating cluttered, 360-degree virtual environments remains a significant challenge due to the limited field of view of standard head-mounted displays, often requiring users to physically scan their surroundings to locate objects. Relying solely on visual search can induce high cognitive load and increase task completion times, necessitating alternative guidance methods.

To address this, we investigated the efficiency of sensory substitution by comparing spatial audio and directional haptic cues for object localization. We conducted a within-subjects experiment ($n = 10$) where participants located target boxes among identical ones using four guidance conditions: visual-only (baseline), spatial audio, directional haptics, and a combined multimodal approach. Directional haptics were delivered via hand controllers, vibrating the left or right controller based on the target's relative direction to the user's gaze. To ensure consistent difficulty, target sequences were pre-defined and counterbalanced rather than purely random.

Our results revealed no statistically significant difference in reaction times between the auditory, haptic, and multimodal conditions, contradicting our initial hypothesis that haptics would be faster. However, a significant effect of the target pattern sequence was observed, suggesting geometry influences search more than the sensory techniques.

Despite equivalent performance, participants expressed a strong subjective preference for haptic feedback, rating it as more intuitive and easier to learn than spatial audio. These findings suggest that directional haptics are a viable and user-preferred alternative to audio for spatial guidance in VR, offering robust utility for accessible design without overwhelming the user.

Video Demo: https://youtube.com/shorts/Pgpqjqmrhi4?feature=share
Repository: https://github.com/tiagoteixeira03/ears-point-to-the-eyes

## Keywords

Virtual Reality, Spatial Audio, Haptics, Multimodal Interaction

## 1 Introduction

Virtual Reality (VR) environments are inherently three-dimensional, offering users a 360-degree field of regard. However, unlike the real world where peripheral vision and auditory cues constantly inform us of our surroundings, VR users are often limited by the specific Field of View (FoV) of the Head-Mounted Display (HMD). This limitation creates significant challenges for target acquisition tasks, particularly when objects of interest are located outside the user's immediate FoV or within cluttered environments where visual scanning induces a high cognitive load [4].

To mitigate these visual limitations, multimodal interaction techniques, specifically sensory substitution, can be employed to guide user attention. By leveraging alternative sensory channels such as audition and touch, developers can provide directional cues that orient the user toward a target without relying solely on visual search.

This study aims to investigate the efficiency of different sensory modalities in aiding object localization in a cluttered 360-degree VR environment. Specifically, we compare the impact of Spatial Audio and Directional Haptics on search reaction time and user preference. We devised a within-subjects experiment where participants were tasked with locating a specific target box among multiple identical distractors arranged on surrounding shelves. The experiment compared four distinct conditions: (1) a baseline visual-only search, (2)

*All three authors contributed equally to this research.

spatial auditory guidance, (3) directional haptic feedback delivered via controllers, and (4) a combined multimodal approach.

Our primary research question investigates whether vibrotactile directional feedback delivered through a hand controller reduces the time required to locate a target compared to spatial auditory cues. Furthermore, we examine user preference regarding ease of learning and comfort, hypothesizing that while haptics may offer intuitive guidance, the combination of audio and haptics might prove overwhelming for some users.

## 2 Related Work

The integration of non-visual cues to enhance spatial awareness and task performance in virtual environments has been a subject of extensive research. This section reviews prior work on auditory localization, haptic guidance, and comparative studies between the two modalities.

### 2.1 Auditory Guidance in VR

Our study is primarily grounded in the work of Hoeg et al. [3], who investigated the effect of audiospatial attention in a 360-degree visual search task. Building on the "pip and pop" effect, where a non-spatial auditory signal temporally synchronized with a visual change facilitates search, Hoeg et al. extended this to VR by comparing no sound, stereo sound, and binaural sound conditions. In their experiment ($n = 10$), participants searched for a horizontal line segment among distractors. Their results indicated that binaural sound, which provides spatial localization cues through Head-Related Transfer Functions (HRTFs), offered a distinct advantage over stereo and no-sound conditions, significantly reducing reaction times. This demonstrates that auditory cues can do more than just alert the user, they can actively guide spatial processing in VR.

Supporting these findings, Larsen et al. [4] ($n = 33$) compared stereo amplitude panning against 3D audio (using non-individualized HRTFs) in a visual search task with varying numbers of distractors. They found that 3D audio reduced search latencies by 28% compared to panning and significantly outperformed the no-audio condition. Crucially, the performance gap widened as the number of distractors increased, and spatial audio continued to improve speeds even when targets were already within the user's FoV.

The importance of realistic audio rendering was further highlighted by Roßkopf et al. [6] ($n = 49$), who compared real loudspeakers against various head-tracked binaural auralizations in an audiovisual VR seminar room. Their findings suggest that realistic binaural renderings support reliable sound source localization and enhance social presence compared to standard game-engine baselines, although distance estimation remained slightly less accurate than with real speakers.

### 2.2 Haptic Guidance

While audio is effective, it competes with other auditory information in the environment. Haptic feedback offers a private, additive channel for guidance. Halabi et al. [2] ($n = 24$) compared reaction times to auditory and vibrotactile cues in a driving simulation using both HMDs and CAVE systems. They reported that mean reaction times for audio cues (5.61s) were slightly shorter than for tactile cues (6.05s). However, they noted that due to the small sample size and small variation in reaction times, it could not be concluded definitively that audio is superior to haptics.

The efficacy of haptics is heavily dependent on implementation parameters. Bao et al. [1] studied young ($n = 10$) and older ($n = 13$) adults to quantify how vibration properties affect reaction times. They found that while small changes in frequency or tactor type had minor effects (10–50 ms), body location and cognitive load had much larger impacts (up to 250 ms), with reaction times increasing as the stimulus moved further from the head. This highlights the importance of cue placement and mental workload in haptic system design.

### 2.3 Multimodal and Accessible Interaction

Haptics also play a critical role in accessibility, serving as a substitute for audio. Mirzaei et al. [5] evaluated EarVR ($n = 40$), a vibrotactile ear-mounted device designed to help Deaf and Hard-of-Hearing (DHH) users localize 3D sounds. DHH participants using EarVR completed sound-based search tasks significantly faster (14.7s) than without it (29.6s), achieving performance comparable to hearing participants. In secondary tasks where DHH users initially failed without feedback, the addition of EarVR allowed them to succeed. This strongly supports the viability of haptics as a primary guidance modality when audio is unavailable or insufficient.

### 2.4 Positioning Our Study

Our work synthesizes these findings by directly comparing *Spatial Audio* (as validated by Hoeg et al. and Larsen et al.) against *Directional Haptics* (as explored by Halabi et al. and Mirzaei et al.) in a static, high-clutter 360-degree search task. Unlike Halabi et al., who focused on driving, or Hoeg et al., who focused purely on audio conditions, our study introduces a "Both" condition to investigate the effects of combining these modalities.

## 3 Methods

We designed a within-subjects user study to evaluate the effectiveness of different sensory modalities in aiding object localization in a cluttered VR environment. The study uses a counterbalanced design with two independent variables: the sensory cue provided and the sequence of target positions.

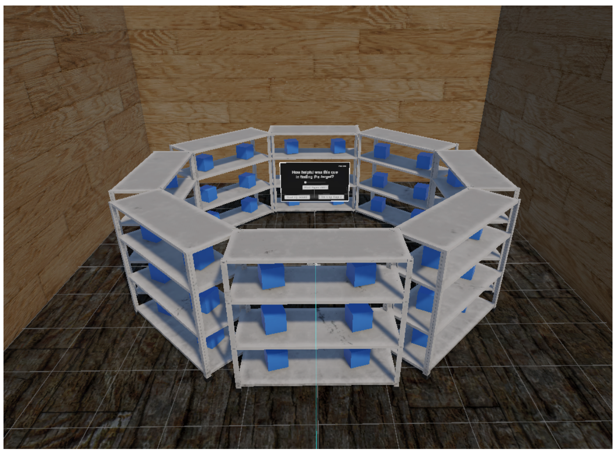
### 3.1 Participants

A total of 10 participants were recruited for the experiment. Before the session, participants completed the profiling questionnaire in Appendix A to collect demographic data, including age range, gender, primary language, and education level. Participants also reported their experience with VR technology and any known visual or auditory impairments. Participation was voluntary, and informed consent was obtained from all subjects.

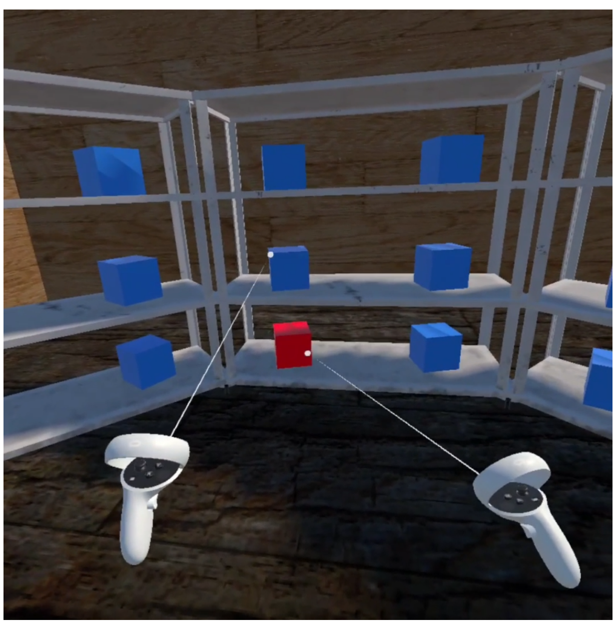
### 3.2 Virtual Environment

The virtual environment was developed using the Unity 6 game engine (version 6000.0.62f1) and the Meta Interaction SDK. The hardware setup included a Meta Quest 2 Head-Mounted Display (HMD) and its accompanying Touch controllers. For development and preliminary testing without the headset, the Meta XR Simulator was utilized.

The virtual scene consisted of a central user position surrounded by eight shelves arranged in a 360-degree radius (Figure 2). Each shelf was populated with six uniform gray cubes, serving as distractors. The target object was a single box that changed color to red upon the start of a trial. The user's task was to locate and select this target box using a raycast pointer emitted from the right hand controller.



(a)



(b)

**Figure 2: The virtual experimental setup: (a) Eight shelves filled with distractors surround a central user position; (b) The user's first-person perspective utilizing the ray interactor to select the active (red) target box.**

## 3.3 Interaction Design and Cues

The experiment implemented four guidance conditions as the primary independent variable:

- **No Cues (Visual Only):** The target box turns red, but no audio or haptic feedback is provided. The user must rely solely on visual scanning.
- **Spatial Audio:** The target box emits a looping "ding" sound. This sound is rendered using Unity's AudioSource with a 3D spatial blend, allowing users to localize the source through audio panning as they rotate their head.
- **Directional Haptics:** The user receives vibrotactile feedback through the hand controllers to indicate the direction of the target.
- **Multimodal (Both):** Both the spatial audio and directional haptic cues are active simultaneously.

*3.3.1 Haptic Feedback Algorithm.* The directional haptics system employs a "hot/cold" navigation logic based on the user's gaze. The system calculates the signed angle between the user's forward vector (head gaze) and the vector pointing to the target box using the following algorithm:

$$\theta = \text{Vector3.SignedAngle}(\vec{F}_{head}, \vec{D}_{target}, \vec{U}_{up}) \tag{1}$$

where $\vec{F}_{head}$ is the forward direction of the headset and $\vec{D}_{target}$ is the direction from the headset to the target.

If the angle $\theta$ is negative (target is to the left), the left controller vibrates. If positive (target is to the right), the right controller vibrates. The intensity of the vibration scales with the magnitude of the angle: stronger vibration indicates a larger angular distance ("turn harder"), while weaker vibration indicates the user is close to the target. A deadzone of $\pm 5°$ is applied where vibration stops to indicate alignment with the target.

## 3.4 Experimental Design and Procedure

To ensure the validity of our results, we introduced a second independent variable: the **Target Sequence**. Randomly selecting a target box for every trial carries the risk of clustering, where a new target appears adjacent to the previous one, artificially lowering search times. To mitigate this, we pre-defined four distinct sequences of 15 target positions (patterns A, B, C, and D). These sequences were designed to offer comparable difficulty and travel distances. For each experimental condition, one of these four sequences was randomly assigned to the participant.

*3.4.1 Protocol.* The experimental session followed a structured protocol:

(1) **Pre-Experiment Form:** Participants filled out the form in Appendix A giving their consent and profiling information.
(2) **Training Phase:** Users entered a training mode to familiarize themselves with the haptic and audio cues. They practiced finding targets until they felt comfortable with the mechanics.
(3) **Main Experiment:** The user completed four blocks of trials, one for each condition (None, Sound, Haptics, Both). The order of conditions was randomized for each participant.

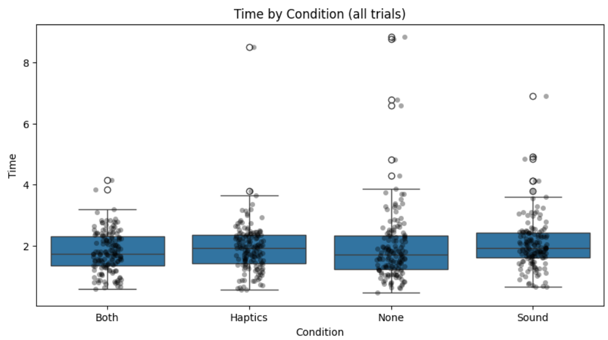



**Figure 3: Boxplot of the reaction times measured for each technique used**

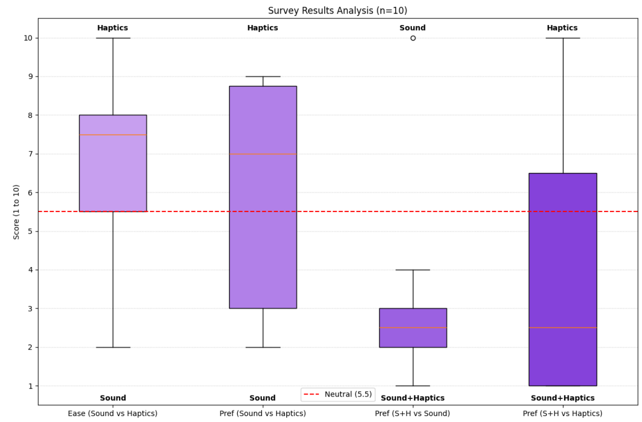



**Figure 4: User's opinion on ease of use and preference for the techniques used.**

(4) **Trials:** Each block consisted of 15 trials. In each trial, a target appeared according to the assigned sequence, and the time taken to select it was recorded.
(5) **In-App Rating:** Immediately after completing a block (15 trials), a menu appeared in VR asking the user to rate the helpfulness of that specific cue on a scale of 1–10.
(6) **Post-Experiment Form:** After completing all four conditions, participants filled out a second form shown in Appendix B where they ranked their preferred techniques.

## 3.5 Hypotheses

Based on the literature and our own testing, we formulated three hypotheses:

- **H1:** Directional haptics will provide the fastest initial orientation but may be less precise for the final selection phase compared to audio.
- **H2:** Directional haptics will be the fastest overall and the most preferred technique by users.
- **H3:** The combination of spatial audio and directional haptics (Multimodal) will be overwhelming and disorienting for most users, leading to lower preference scores.

# 4 Results

## 4.1 Reaction Time

Figure 3 presents the reaction times measured for each technique.

For the statistical analysis, we first assessed whether the relationship between the techniques used and the collected reaction times could be evaluated using a two-way ANOVA. To this end, a Shapiro–Wilk test was applied to all treatments of both independent variables. The results indicate that the data follow a normal distribution, as all corresponding p-values were below 0.05.

We then conducted a two-way ANOVA with technique and pattern as independent variables and reaction time as the dependent variable, in order to evaluate their statistical significance.

The ANOVA did not reveal a statistically significant effect of the technique used ($p = 0.256$) therefore, hypotheses H1 and H2 cannot be supported. These results are illustrated in Figure 3.

In contrast, the ANOVA revealed a statistically significant effect of the pattern used ($p = 0.017$). Although this finding does not directly confirm any of the proposed hypotheses, it provides relevant insight for the discussion of the results.

To identify which patterns contributed to this significant effect, a Tukey post-hoc test was performed. This analysis revealed a statistically significant difference only between `PatternA` and `PatternB` ($p = 0.012$), with a mean difference of 0.327 s. This indicates that `PatternA` is, on average, 0.327 seconds slower than `PatternB`.

## 4.2 User Preference

Figure 4 presents the user preferences collected through the post-experiment questionnaire. The majority of participants reported that haptics were easier to learn compared to sound. When asked to indicate their preference between haptics and sound, users also expressed a preference for haptics. Finally, when comparing each technique individually with the simultaneous use of both techniques, participants reported a strong preference for the combined approach.

These results do not support H3, as the observed preference indicates that users did not feel overwhelmed or uncomfortable when using the combination of both methods.

# 5 Discussion

## 5.1 Technique Performance

Unfortunately, we cannot draw conclusions regarding the effects of auditory and tactile cues on improving users' spatial awareness in a VR environment, as the obtained results did not provide sufficient evidence to support any of the proposed hypotheses.

As similar effects have been successfully demonstrated in [3], we attribute this outcome to limitations in our testing methodology.

The lack of statistical significance for the techniques used can be attributed to three likely factors. First, the reduced sample size limited the reliability of the results. With only 10 participants, it is difficult to obtain consistent findings that adequately support the hypotheses. Second, the number of trials conducted per technique was relatively small. Given that the average testing time per participant was approximately 10 minutes (including questionnaire completion), the number of data points collected per technique could have been increased to 20 or even 30 without significantly increasing user fatigue. Finally, the patterns used may have influenced the observed outcomes. Although the patterns were designed to be equally difficult, this was not the case, as `PatternA` was shown to be significantly more difficult than `PatternB`.

We also identified an additional observation that may help explain the obtained results. Among all tested techniques, the condition without any cues produced the largest number of outliers. Further analysis revealed that most of these outliers occurred during a specific scenario: the first trial without cues. We suspect this occurred because participants were not aware that the test had begun, as no explicit indication was provided. This issue would likely have been identified and addressed through the inclusion of a pilot study, during which participants could have reported this ambiguity.

Considering these factors, several improvements can be identified for a future iteration of the experiment. The study would include a larger sample size and collect a greater number of data points per technique. Additional care would be taken to ensure that the patterns used are equivalent in difficulty, so as not to confound the results. Furthermore, the experimental setup would be modified to clearly indicate when a trial begins, either by explicitly informing participants beforehand or by displaying the active technique within the testing interface. Finally, a pilot study would be conducted to identify and mitigate potential issues prior to the main experiment.

### 5.2 User Preference

The results indicate a clear preference for haptics in both the learning process and overall usability. Although this preference does not translate into improved performance in terms of spatial awareness, it remains an important consideration when discussing effective methods for conveying cues in a three-dimensional space.

We also conclude that users did not perceive the combination of haptic and auditory cues as overwhelming and, in fact, preferred it over the use of a single cue.

## 6 Conclusion

In this study, we investigated the efficiency of multimodal sensory substitution techniques for object localization in a cluttered, 360-degree Virtual Reality environment. Specifically, we compared the impact of spatial audio, directional haptic feedback, and a combined multimodal approach against a visual-only baseline on search reaction times and user preference. Our goal was to determine if vibrotactile directional cues delivered through hand controllers could offer a faster or more intuitive alternative to traditional spatial audio cues.

Contrary to our initial hypotheses, the quantitative results did not reveal a statistically significant difference in reaction times between the four techniques ($p > 0.05$). This suggests that, within the constraints of our specific experimental setup and sample size ($n = 10$), neither modality provided a definitive performance advantage over the others. However, we did observe a significant effect of the target pattern sequence ($p < 0.05$), indicating that the geometric distribution of targets played a larger role in search efficiency than the sensory modality itself.

Despite the lack of objective performance differentiation, the qualitative findings were robust. Participants expressed a strong preference for directional haptics, rating it as easier to learn and more intuitive than spatial audio. Furthermore, the multimodal condition (combining audio and haptics) was the most preferred overall, contradicting our hypothesis that it might be overwhelming. Users did not report disorientation but rather appreciated the redundant cues.

These findings highlight the potential of haptic feedback as a viable and user-preferred alternative for spatial guidance in VR, particularly for accessibility or scenarios where audio is unavailable. For future work, we recommend:

(1) **Increasing the sample size** and the number of trials per condition to improve statistical power and mitigate the impact of outliers.
(2) **Conducting a pilot study** to refine the experimental protocol, specifically to ensure that start-of-trial indicators are clear to preventing "false" outliers in the baseline condition.
(3) **Balancing target sequences** more rigorously to ensure consistent difficulty across all experimental conditions, potentially by using algorithmically generated equidistant paths rather than pre-defined patterns.
(4) **Investigating different haptic encodings**, such as varying frequency or temporal patterns, to see if more complex tactile cues can outperform simple directional vibration.

## References

[1] Tian Bao, L Su, C Kinnaird, M Kabeto, PB Shull, and KH Sienko. 2019. Vibrotactile display design: Quantifying the importance of age and various factors on reaction times. *PLOS ONE* 14, 8 (2019), e0219737.
[2] Osama Halabi, M Bahameish, L Al-Naimi, and A Al-Kaabi. 2019. Response Times for Auditory and Vibrotactile Directional Cues in Different Immersive Displays. *International Journal of Human-Computer Interaction* 35, 17 (2019), 1578–1585.
[3] Emil R. Hoeg, Lynda J. Gerry, Lui Thomsen, Niels C. Nilsson, and Stefania Serafin. 2017. Binaural sound reduces reaction time in a virtual reality search task. In *2017 IEEE 3rd VR Workshop on Sonic Interactions for Virtual Environments (SIVE)*. 1–4. doi:10.1109/SIVE.2017.7901610
[4] C Larsen, D Lauritsen, J Larsen, M Pilgaard, and J Madsen. 2014. Aurally Aided Visual Search Performance Comparing Virtual Audio Systems. In *Audio Engineering Society Convention 137*. Audio Engineering Society.
[5] M Mirzaei, P Kán, and H Kaufmann. 2020. EarVR: Using Ear Haptics in Virtual Reality for Deaf and Hard-of-Hearing People. *IEEE Transactions on Visualization and Computer Graphics* 26, 5 (2020), 2084–2093.
[6] S Roßkopf, LOH Kroczek, F Stärz, M Blau, S Van De Par, and A Mühlberger. 2024. The impact of binaural auralizations on sound source localization and social presence in audiovisual virtual reality. *Acta Acustica* 8 (2024), 72.

## A Pre-experiment Form

# Testing sound and haptics in a virtual reality search task

Collection of essential information about the subject before a user experiment, including demographics, handedness, experience and perceptual conditions.

**renou.tiago@gmail.com** Mudar de conta

Não partilhado

* Indica uma pergunta obrigatória

I understand that the participation in this study is voluntary and that I may withdraw at any time without providing any explanation. If this happens, I will not be subject to any penalty and the data relating to my experience will be removed and destroyed. *

- ◯ I agree.

User ID *

A sua resposta

Age Range *

- ◯ 18-25
- ◯ 26-35
- ◯ 36-50
- ◯ more than 50

Primary Language *

- ○ Portuguese
- ○ English
- ○ Outra:

Highest Degree or Level of School Completed *

- ○ No schooling completed
- ○ High school graduate
- ○ Trade/technical/vocational training
- ○ Bachelor's degree
- ○ Master's degree
- ○ Professional degree
- ○ Doctorate degree

Gender/Sex *

- ○ Feminine/Female
- ○ Masculine/Male
- ○ Other

How often do you use VR? *

- Never
- Rarely (e.g., once a month or less)
- Occasionally (e.g., a few times a month)
- Frequently (e.g., once a week or more)
- Very Frequently (e.g., daily)

Did you perform a previous version of the current setup or do you have any knowledge about the test and protocol (Naivety to the Protocol)? *

- Yes
- No

How do you classify your vision? *

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Blindness | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Perfect |

Do you have any other vision problem? (e.g., Myopia, Hyperopia, Color Blindness, Astigmatism)

A sua resposta

How do you classify your hearing quality? *

- Normal (less than 15dB)
- Slight (16dB to 25dB)
- Mild (26dB to 40dB)
- Moderate (41dB to 55dB)
- Moderately Severe (56dB to 70dB)
- Severe (71dB to 90dB)
- Profound / Deaf (more than 91dB)

Do you have any different characteristic on the skin of the hands (e.g. Loss of sensibility, Scar, Hyperhidrosis)?

A sua resposta

Seguinte

Limpar formulário

## B Post-experiment Form

# Testing sound and haptics in a virtual reality search task

**renou.tiago@gmail.com** Mudar de conta

Não partilhado

* Indica uma pergunta obrigatória

Post-Experiment Feedback

Which technique was easier to learn? *

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Sound | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Haptics |

Which technique did you prefer? *

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Sound | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Haptics |

Which technique did you prefer? *

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Sound + Haptics | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Sound |

Which technique did you prefer? *

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Sound + Haptics | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Haptics |

Anterior Enviar Limpar formulário