Estimating the Effects of Ambient Noise on Mixed Reality Interaction

TINGHUI LI*, University of Sydney, Australia
DANYANG PENG, Keio University Graduate School of Media Design, Japan
EDUARDO VELLOSO, University of Sydney, Australia
ANUSHA WITHANA, University of Sydney, Australia
KOUTA MINAMIZAWA, Keio University Graduate School of Media Design, Japan
ZHANNA SARSENBAYEVA, University of Sydney, Australia

Mixed Reality devices, now as ubiquitous as smartphones and laptops, are increasingly utilised in everyday scenarios. However, real-world use often involves challenging environmental factors such as ambient noise, which can adversely affect user interaction with MR systems. This study investigates the impact of three types of ambient noise—music, urban noise and speech—on MR interaction. We constructed Bayesian regression models to assess movement time, pointing offset, error rate and throughput on target acquisition task, and throughput, uncorrected error rate, corrected error rate, and words per minute on text entry task under different noise conditions. Our results indicated that meaningless speech reduced text-entry throughput by 5.36%, fast-tempo music increased movement time by 4.44%, slow-tempo music increased pointing offset by 4.07%, urban indoor noise increased typing throughput by 3.33% and urban outdoor noise decreased throughput by 2.74%. These findings demonstrate how ambient noise affects MR performance, advancing our understanding of situational impairments in MR. We propose strategies for designing noise-resilient MR interfaces to enhance usability in dynamic environments.

CCS Concepts: • **Human-centered computing → Mixed / augmented reality**; **Empirical studies in HCI**; **Empirical studies in accessibility**.

Additional Key Words and Phrases: Situational Impairments, Ambient Noise, Fitts's Law, Text Entry

ACM Reference Format:
Tinghui Li, Danyang Peng, Eduardo Velloso, Anusha Withana, Kouta Minamizawa, and Zhanna Sarsenbayeva. 2025. Estimating the Effects of Ambient Noise on Mixed Reality Interaction. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 9, 4, Article 189 (December 2025), 34 pages. https://doi.org/10.1145/3770715

1 INTRODUCTION

Mixed Reality (MR) devices are becoming as ubiquitous as smartphones and laptops, with early adopters already showcasing their potential in everyday scenarios such as work[1] and shopping[2]. For instance, users can wear headsets in office environments to perform tasks like document editing and responding to emails, often while

*Corresponding author.
[1]https://youtu.be/0iV--jRWBig?si=aLlWrk7BWcIZQxea [Accessed: 2025-01-24]
[2]https://youtu.be/a0EZPyUjJQc?si=Irkum0vUnAMbL0YO [Accessed: 2025-01-24]

Authors' Contact Information: Tinghui Li, University of Sydney, Sydney, Australia, tinghui.li@sydney.edu.au; Danyang Peng, Keio University Graduate School of Media Design, Yokohama, Japan, pengdanyang@keio.jp; Eduardo Velloso, University of Sydney, Sydney, Australia, eduardo.velloso@sydney.edu.au; Anusha Withana, University of Sydney, Sydney, Australia, anusha.withana@sydney.edu.au; Kouta Minamizawa, Keio University Graduate School of Media Design, Yokohama, Japan, kouta@kmd.keio.ac.jp; Zhanna Sarsenbayeva, University of Sydney, Sydney, Australia, zhanna.sarsenbayeva@sydney.edu.au.

ACM 2474-9567/2025/12-ART189
https://doi.org/10.1145/3770715

listening to music or conversing with colleagues. Similarly, shoppers can leverage MR headsets for immersive product exploration, such as virtually trying on outfits or accessing product details in noisy shopping malls while engaging with friends. These real-world scenarios often expose users to non-conventional contextual and environmental factors that can negatively impact their interaction with MR systems, leading to *situational impairments* [66]. One critical factor is ambient noise—the background sounds that users are exposed to while focusing on specific tasks or activities [26, 63], and therefore, we aim to investigate the effects of different ambient noise on MR interaction.

Understanding the impact of contextual factors on system interaction is the goal of recent research in ubiquitous computing [60]. Prior research has shown that environmental and contextual factors, such as ambient noise [63], stress [60], ambient light [61], encumbrance [36, 51], and walking [22, 36], negatively affect user interaction on various devices, including desktop computers [56], smartphones [58, 60, 63], wearable displays [23, 87], and MR devices [36, 37]. For instance, walking or carrying objects has been found to impair more than 50% of user performance with MR devices [36]. Additionally, ambient noise has been shown to significantly hinder smartphone usage [63], where the urban outdoor noise and meaningful speech had a significant impact on participants' performance during typing. This has expanded the field of research in Human-Computer Interaction (HCI) dedicated to designing systems that can identify situational impairments in users and adjust interfaces accordingly to reduce their impact on interaction [64]. Therefore, understanding and modelling the consequences of these impacts is the primary step toward sensing and adapting these effects that cater to real-world scenarios and ensure usability in diverse and challenging environments [59, 73, 78].

Recent research in cognitive science and HCI has demonstrated that environmental noise can adversely impact user performance [63], alter human behavior [24, 54], and affect emotional well-being [26]. This negative influence is primarily attributed to the dispersion of attention during task execution [5]. Even when users intentionally direct their focus elsewhere, ambient sounds are still processed by the brain, resulting in diverted attention and reduced performance [5, 63]. In fact, Sarsenbayeva et al. [63] have empirically shown that ambient noise impairs mobile interactions under certain conditions. Motivated by these findings, we investigated whether ambient noise similarly affects user performance in MR interactions and how its impact compares to other situational impairments evaluated within MR environments.

In this paper, we investigate the effects of ambient noise—a situational impairment that remains relatively underexplored—on MR interactions. We simulated three types of ambient noise: music, urban noise, and speech, with two variants of each. Participants performed tasks commonly used in Fitts' Law experiments in MR, such as 2D target acquisition via direct selection and ray-casting, as well as text entry tasks [36]. We incorporated Bayesian regression models to measure the effects of noise on movement time [36, 51], pointing offset [36, 61], and error rate [25] in the target acquisition tasks, and throughput [84] in the text entry task. Our results indicated that meaningless speech decreased text entry throughput by 5.36%, fast-tempo music increased movement time by 4.44%, slow-tempo music increased offset by 4.07%, urban indoor noise increased typing throughput by 3.33%, and urban outdoor noise decreased throughput by 2.74%. The target acquisition throughput decreased substantially under all ambient noise conditions except meaningless speech and fast-tempo music condition. Besides, meaningless speech, fast-tempo music, and slow-tempo music conditions decreased text entry throughput.

In summary, this study advances the expanding research field on situational impairments in MR by addressing the impact of ambient noise during MR interactions. The key contributions of this paper are as follows:

- We build Bayesian regression models to quantify by how much ambient noise increase movement time and decrease throughput;
- We enhance the understanding of the effects of ambient noise on MR interaction and contribute towards accumulating knowledge in situational impairments research expanding it to MR;

- We highlight the importance of considering situational impairments in MR interaction and propose potential strategies and directions to mitigate the impacts of situational impairments.

2 RELATED WORK

2.1 Situational Impairments on Mixed Reality Interaction

With the advancement of MR devices, their usage is increasingly comparable to that of smartphones and laptops, offering a more immersive experience than traditional devices. As MR becomes more ubiquitous, it inevitably encounters non-conventional contextual and environmental challenges [35]. These challenges are referred to in the literature as Situationally-Induced Impairments and Disabilities (SIIDs), or situational impairments for short [66]. Unlike health-related impairments, SIIDs are caused by external environmental factors rather than an individual's physical or mental health [78]. Once the external factors are resolved, the situational impairments disappear, restoring the user's normal capabilities [33, 34, 38].

To seamlessly integrate MR technologies into daily tasks, it is essential to understand how situational impairments influence MR interactions [39]. As technological interactions occur in diverse contexts, these impairments are inevitable [78]. Previous studies have shown that situational impairments significantly impact the performance of mobile devices. While their effects on mobile devices [62, 73], desktop computers [56], and wearable devices [23] have been studied extensively, their impact on MR headset interactions remains underexplored. To address this gap, our study contributes to the growing body of situational impairment research by quantifying the effect of noise on MR interaction and measuring changes in user performance during canonical MR tasks.

Previous research has examined the effects of encumbrance and walking on MR interaction [36]. Findings indicate that a 1.0 kg encumbrance increases selection movement time by 28% and decreases text entry throughput by 17%. Walking leads to a 63% increase in ray-casting movement time and a 51% reduction in text entry throughput. Additionally, walking increases direct selection pointing offset by 16%, ray-casting pointing offset by 17%, and error rate by 8.4%. These findings underscore that situational impairments have a substantial impact, warranting further investigation. In this paper, we extend our research by comparing the effects of ambient noise with those of encumbrance and walking. By examining these varied conditions, we aim to understand how each impairment uniquely influences user performance and to inform the development of more resilient interaction systems.

2.2 Ambient Noise as Situational Impairment

Augmented Reality (AR) acoustics focuses on the real-time synthesis and rendering of spatialised audio that coexists with or augments a user's physical environment [49]. Early approaches relied on head-related transfer functions for virtual sound placement, with adaptive filtering to account for dynamic listener movement and environmental changes [6, 86]. More recent systems leverage room acoustic modelling to deliver more realistic reverberation and occlusion effects in AR settings [82]. While AR acoustics research has largely concentrated on refining algorithms for reverberation, occlusion, and real-time reflections, our study shifts the lens toward how everyday ambient noise degrades user perception and task performance in MR settings. By treating ambient noise as an external situational impairment, we aim to understand and mitigate its impact on MR performance.

Previous studies have highlighted how ambient noise impacts human behavior [29, 46, 63], emotions [26], and cognitive performance [4, 76, 81]. Sarsenbayeva et al. [63] split the common noise into three categories: music, ambient noise, and speech. Their results showed that both fast and slow music can reduce the time to complete a task [63]. Similarly, Woo Ee and Kanachi [79] found that classical music improved recall performance compared to rock music. Milliman [50] showed that the music with fast tempo increased the walking speed of customers. Holbrook [24] demonstrated that fast-tempo music requires greater cognitive effort to process compared to slow-tempo music. North et al. [54] explained this by noting that the listener's brain perceives and processes more information when exposed to fast-tempo music than to slow-tempo music. Therefore, in our study, we evaluated the effects of music with fast and slow tempo that except lyrics.

Another category of noise introduced by Sarsenbayeva et al. [63] is urban ambient noise. Several studies have demonstrated the negative impact of urban outdoor noise on human performance. Stansfeld et al. [71] found that communities exposed to lower levels of traffic noise experienced reduced rates of psychiatric hospitalizations. Additionally, Arnsten and Goldman-Rakic [1] revealed that exposure to broadband noise can significantly impair performance on memory tasks. Similar to outdoor urban noise, Banbury and Berry [4] showed that the presence of urban indoor noise significantly hindered students' performance on mathematical and memory recall tasks. However, Sarsenbayeva et al. [63] revealed a mixed impact of urban noise: while participants performed faster on target acquisition and memorization tasks, likely motivated by a desire to escape the unpleasant noise, they perceived the noise as highly distracting. Therefore, we evaluated the effects of urban noise with outdoor and indoor scenarios.

Studies have distinguished between the cognitive effects of meaningful and meaningless speech on human cognition [45]. For example, Martin et al. [45] discovered that continuous meaningful speech had a negative impact on reading performance compared to silence, with reading comprehension suffering under both meaningful and meaningless speech, though the former had a stronger impact. The impact of speech was observed only in the text entry task, where participants took significantly longer to type while listening to meaningful speech [63]. This aligns with previous research showing that meaningful speech negatively affects cognitive performance. Other studies have shown that meaningless speech negatively impacts memory tasks. Salamé and Baddeley [57] demonstrated that unattended foreign language speech caused the greatest disruption to immediate memory compared to other conditions, such as instrumental music, urban noise, and silence. Sarsenbayeva et al. [63] noted that listening to understandable English speech led to slower completion times, while the effect of meaningless Kazakh speech was much smaller, likely because it was easier to ignore. Therefore, in our study, we evaluated the effects of speech with meaningful and meaningless types.

2.3 Common MR Tasks for Performance Evaluation

In recent MR evaluation studies, researchers have converged on a set of interaction tasks that closely mirror the natural hand-based gestures we perform in everyday life, such as tapping to select, pointing to indicate, and typing to communicate. Drawing on this [18, 36, 83], we implemented three core tasks in our user study: target acquisition via direct selection, target acquisition via ray-casting, and text entry. To eliminate potential confounding factors and ensure clarity in assessing the effects of noise on MR interaction, we selected and utilised these core simple tasks that avoid the additional cognitive load or processes introduced by more complex tasks [72]. Nevertheless, the chosen tasks still draw on established MR paradigms for studying interaction under real-world constraints. By having participants manipulate virtual objects while still seeing the real world surroundings, it mirrors the way people use MR, performing reach-and-point, grab-and-place, and pinch gestures that are commonly used in commercial headsets.

Target acquisition (direct selection) has become a standard benchmark for evaluating MR pointing performance and motor control under varying difficulty levels. Studies by Li et al. [36] and Yu et al. [83] all employ a Fitts' ring paradigm in which eleven targets are evenly spaced on a circular layout. Following Meta's hand-interaction guidelines, the ring is positioned approximately 49 cm from the headset's central lens and comprises three target diameters (4 cm, 6 cm, 8 cm) and two inter-target distances (3 cm, 5 cm), yielding six configurations that systematically vary the index of difficulty (ID). This well-validated protocol supports rigorous cross-study comparisons of movement time and error-rate metrics.

In addition to direct selection, ray-casting has emerged as a prevalent selection technique in MR research, leveraging a virtual "laser" to extend user reach and improve precision at a distance. Prior work by Li et al.

[36] and Zhou et al. [85] has demonstrated how Fitts' ring layouts can be adapted for ray-casting, showing comparable throughput and error-rate trends to direct selection while offering ergonomic benefits. By including this interaction paradigm alongside tap-based tasks, researchers can more fully characterize how different input modalities respond to factors like situational impairments or environmental complexity.

Prior MR interaction studies on situational impairments consistently evaluate pointing performance using established metrics, including movement time (MT), pointing offset, and error rate [36, 60, 63]. MT quantifies how long users take to acquire a target as a function of its size and distance. Pointing offset measures the spatial distance between the selection endpoint and the target center, capturing precision. Error rate reflects the proportion of incorrect or missed selections, indicating overall reliability. By situating our direct-selection task within this framework, we align our methods with a rich body of MR benchmarks and enable direct comparison of how ambient noise impairs these core performance metrics.

Text entry tasks in MR frequently draw upon desktop and mobile typing paradigms to evaluate input performance under varying situational factors. Prior investigations by Dube and Arif [18], Speicher et al. [70], and Li et al. [36] have all adapted MacKenzie's phrase set[43]—a standard collection of sentences—to ensure consistent lexical and syntactic complexity across conditions. These studies highlight how error-rate metrics, including uncorrected error rate (UER) and corrected error rate (CER), can reveal subtle degradations in speed-accuracy trade-offs when users contend with impairments such as audio distractions. By aligning our text-entry evaluation with this body of work, we ground our throughput and error analyses in well-validated benchmarks.

Throughput, which jointly captures speed and accuracy, has long served as a unifying performance metric in text-entry studies across desktop, mobile, and MR interfaces [13, 77]. Early models simply divided characters entered by time, but more recent work incorporates error penalties to reflect real-world typing conditions [17]. Zhang et al. [84] proposed an independent throughput metric based on Shannon information theory [67] that takes into account both UER and CER, and character per second (CPS) entry rate [68, 84]. This approach has been validated in MR text-entry tasks under varied situational impairments, demonstrating its sensitivity to subtle speed-accuracy trade-offs.

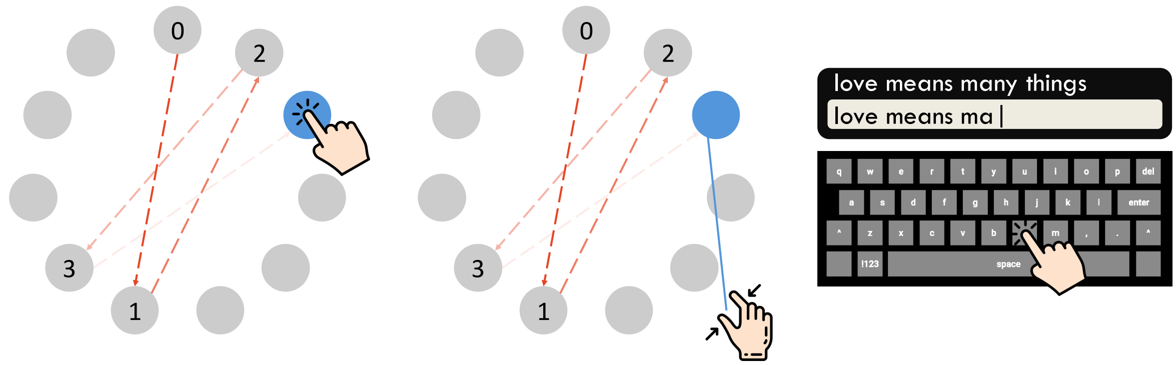
3 METHOD

In this study, we investigate the effect of noise on three common MR interaction tasks followed by previous studies on situational impairments [36, 52, 63]: target acquisition via direct selection, target acquisition via ray-casting, and text entry. We conducted a within-subjects design and simulated different noise types in a controlled environment to exclude other factors that might potentially interfere with the experimental results. The study received ethical approval from the Keio University Graduate School of Media Design Ethics Committee and the Human Research Ethics Committee of the University of Sydney (Application No. 2019/553).

3.1 Tasks

We designed three tasks that closely follows the design from the evaluation of other SIIDs on MR interaction [36]. We developed a custom MR application for the Meta Quest 3 in Unity `2022.3.12f1` and presented these tasks to participants in a counterbalanced order to avoid sequence effects. We describe each task in detail below.

3.1.1 Target Acquisition – Direct Selection. In this task, participants were instructed to select the blue target by tapping it with the index finger of their dominant hand. Eleven targets were evenly distributed along the circumference of a circle (Figure 1a). Different IDs were presented in a random order. Participants completed six repetitions per block, with time-stamped logs capturing selection timing and spatial coordinates for subsequent movement-time and accuracy analyses.



(a) Target Acquisition (Direct Selection) (b) Target Acquisition (Ray-Casting) (c) Text Entry

Fig. 1. The visual representation of the tasks used in the experiment: (a) target acquisition (direct selection); (b) target acquisition (ray-casting); (c) text entry. Blue circle represents the target that needs to be selected. Participants follow the path indicated by the red dashed arrows to select the targets sequentially.

3.1.2 Target Acquisition – Ray-Casting. Participants performed the ray-casting version of the Fitts' ring task by forming a pinch gesture to emit a visible line from their hand and "shoot" at targets that alternated in blue as cues (Figure 1b). Each participant completed eleven selections per configuration, cycling through all six in a counterbalanced order. Trial logs captured selection timestamps and 3D coordinates to compute movement time and accuracy metrics.

3.1.3 Measures of Target Acquisition. Building on previous studies on SIIDs [36, 60, 63], we evaluated performance in target acquisition tasks using metrics such as movement time, pointing offset, and error rate.

Movement time (MT) refers to the duration (in milliseconds) participants take to select a target [20]. It quantifies the time required to move to a target area as a function of the distance to the target and the size of the target [20]. MT is linearly associated to the index of difficulty (ID) [41], as expressed in Equation 1. The ID is determined by the target width (W) and distance (D) [41], as shown in Equation 2, where a and b are empirically derived coefficients. Lower movement times indicate better performance.

$$MT = a + b \cdot ID \tag{1}$$

$$ID = \log_2 \left(\frac{D}{W} + 1 \right) \tag{2}$$

To assess the precision of target acquisition, we used pointing offset as a metric. This corresponds to the distance between the end-point location (i.e., where the user selects or reaches) and the centre of the target [51, 60]. As such, it is an error measure, quantifying how close the user's final action is to the intended target's centre. Pointing offset was calculated as the distance between the target center (x_0, y_0, z_0) and the actual user selection location $(\hat{x}, \hat{y}, \hat{z})$, as defined in Equation 3. Higher pointing offset values indicate lower performance accuracy.

$$\text{Offset} = \sqrt{(\hat{x} - x_0)^2 + (\hat{y} - y_0)^2 + (\hat{z} - z_0)^2} \tag{3}$$

Different from pointing offset, error rate measures the frequency of mistakes made during target selection [60]. It was calculated as the percentage of incorrectly selected targets relative to the total number of targets, as defined in Equation 4. This includes uncorrected selections, wrong selections, failed attempts to engage with virtual elements or misinterpretations of interface prompts. A lower error rate suggests that the user can effectively interact with the MR system [36].

$$E_{target} = \frac{\text{Incorrect Selections}}{\text{Total Targets}} \cdot 100\% \tag{4}$$

3.1.4 Text Entry. Participants typed five randomly selected sentences from MacKenzie's set [43] into a virtual text box, copying each phrase exactly as shown. They were free to submit with or without corrections, allowing us to capture both UER (the proportion of uncorrected mistakes) and CER (the proportion of mistakes that were corrected) for a comprehensive throughput calculation [68, 84]. Sessions were counterbalanced across noise conditions.

3.1.5 Measures of Text Entry. We adopt the independent throughput metric of Zhang et al. [84] to evaluate our text-entry task. The calculation is shown in Equation 5, where $I(X, Y)$ is the transmitted information that can be calculated as $I(X, Y) = H(X) - H_Y(X)$, with $H(X)$ representing the source information, and $H_Y(X)$ representing the conditional entropy.

$$\text{Throughput} = I(X, Y) \cdot CPS \tag{5}$$

We used both UER and CER to measure the performance in the text entry task.The UER represents the number of mistyped characters that remain uncorrected during entry (Equation 6), whereas the CER captures the mistakes that participants successfully rectify (Equation 7).

$$UER = \frac{\text{Uncorrected and Not Fixed Character}}{\text{Total Character}} \cdot 100\% \tag{6}$$

$$CER = \frac{\text{Uncorrected but Fixed Character}}{\text{Total Character} + \text{Uncorrected but Fixed Character}} \cdot 100\% \tag{7}$$

3.2 Conditions

The experimental noise conditions chosen for this study were based on their proven effects on human behavior and cognitive performance, as reported in the literature and evaluated in previous studies:

- Silence: control group
- Music: fast and slow tempo
- Urban ambient noise: indoor and outdoor
- Speech: meaningful (English/Chinese/Japanese) and meaningless (Kazakh)

We used the same musical composition at both fast and slow tempos for music condition to ensure that other musical elements like pitch and timbre remained constant [63]. We avoided using music with lyrics to eliminate potential interference with the speech condition [26]. Following these criteria and previous study design [63], we selected Bach's Brandenburg Concerto No. 2[3] Allegro version for the fast tempo condition. For the slow tempo condition, we used the same music but in its Andante version, which is below the 70 beats per minute threshold commonly defined as slow in the literature [26].

[3]https://youtu.be/8z-l5sF8Vp8?si=01vr7l0VVSYwKgIc [Accessed: 2025-01-24]

We chose a sounds of the city audio clip[4]—including traffic, street noise, the sound of cars, cable cars, and crowd noises—for the urban outdoor ambient noise condition. We selected a coffee shop background noise[5]—including sounds from a coffee machine, cutlery, and indistinct conversations—for the urban indoor ambient noise condition.

For the meaningful and meaningless speech condition, we used a weather forecast script obtained online[6], which is originally displayed in English. The script was translated into Chinese and Japanese for the meaningful condition, accommodating speakers of different native languages. In this way, we played distinct audio tracks tailored to each language group to ensure that all participants can directly comprehend the content. For the meaningless condition, it was translated into Kazakh—a language unfamiliar to all participants. The script was delivered using AI-generated speech[7]. All clips contained only the speaker's voice, with no additional sounds.

The sound was provided through headphones to minimise interference from other environmental sounds. We used Sony MDR-CD900ST monitor headphones, which are professional-grade, closed-back headphones without active noise cancellation. However, we kept the room quiet to avoid disturbing the participants. The playback volume was consistently maintained at approximately 75% of the system's maximum output to ensure clarity while avoiding discomfort or auditory fatigue. We computed the integrated loudness for every stimulus using an online loudness meter[8]. The integrated loudness of the speech stimuli lies in the range −32.6 to −30.6 LUFS, and the music stimuli in the range −18.7 to −14.9 LUFS. While these levels are somewhat quieter than certain broadcast or streaming targets for speech (often around −23 to −18 LUFS), they are still within a plausible and commonly reported range in psychoacoustic/linguistic experiments and allow for consistent comparisons among stimulus types [32]. The urban ambient noises are within an acceptable range, with outdoor ambient noise louder than the indoor condition. Table 4 in Appendix A shows the detailed measurement and all audio resources are attached in the supplementary. Part of the audio clips were sourced from YouTube. All audio clips can be found in the supplementary material. An example of the experimental environment is shown in Figure 2.

Fig. 2. Representative images from the user study sessions.

[4]https://youtu.be/eXHURaIl7hA?si=HeVqfgZi-SPoFtR4 [Accessed: 2025-01-24]
[5]https://youtu.be/BOdLmxy06H0?si=6fYNgGC66xGpvqaS [Accessed: 2025-01-24]
[6]http://www.putlearningfirst.com/language/09trans/weather.html [Accessed: 2025-01-24]
[7]https://elevenlabs.io/ [Accessed: 2025-01-24]
[8]https://youlean.co/file-loudness-meter/ [Accessed: 2025-09-19]

3.3 Participant

We recruited 30 participants (11 men, 18 women, 1 other) in our study using our university's notice board and snowball recruitment. Participants' average age was 26.6 years (min = 22, max = 34, SD = 2.93). Among the participants, 28 were right-handed and 2 were left-handed. 30 out of 30 participants received their bachelor's degree or higher (17 Master's, 11 Bachelor's, 2 Doctorates). The foreign language presented in this study was not known to any of the participants. 20 out of 30 participants were not familiar with the headset. Each participant was given a unique anonymous ID in our study. Our sample size is along the sample size standards for HCI research [10].

3.4 Procedure

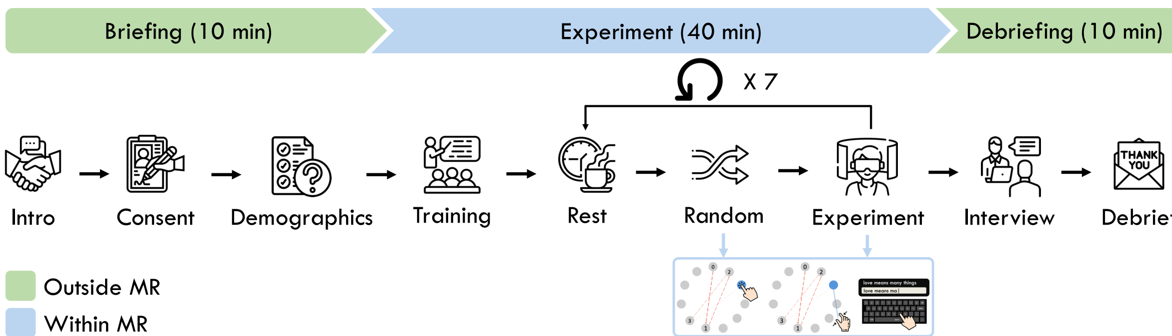


Fig. 3. The experimental procedure.

Our study followed a within-subjects design with different conditions acted as independent variables. The order of conditions was randomized for each participant to minimise potential fatigue or learning effects. Upon arrival at the lab, participants were given an overview of the study's purpose. After confirming their understanding and willingness to participate, they signed a consent form. Participants were then asked to complete a background questionnaire, which gathered demographic information such as gender, age, dominant hand, education level, native language, background, and experience with MR. This data ensured our participant pool was diverse and representative.

We trained our participants using a tutorial that covered all three tasks, ensuring they became comfortable and familiar with both the headset and the tasks to minimise learning effects. The tutorial featured simplified versions of the actual tasks: the target acquisition tasks (direct selection and ray-casting) included only a single trial of the Fitts' ring, while the text entry task involved typing a single sentence not included in the set of phrases used during the study.

After completing the training, we began video recording, which continued until the conclusion of the interview. Participants performed the three tasks under the researcher's guidance. To prevent fatigue and dizziness, participants were allowed a rest period of 1–2 minutes between each condition, adjusted based on their perception to ensure they felt adequately rested. At the end of the experiment, we conducted a semi-structured interview with each participant to gather additional insights into their experience. For the semi-structured interviews, The discussions centred on participants' subjective experiences and reflections while performing tasks under seven distinct ambient sound conditions. To guide the interviews, we focused on the following questions: "Do you typically listen to any kind of sound while working on tasks?", "During these MR tasks, which type of sound do you like or dislike the most?", and "Do you think different sounds affect your mood or task performance in

MR environments?". Follow-up questions were asked based on participants' individual responses to facilitate deeper discussion. All interviews were conducted in English, audio-recorded, and subsequently transcribed and manually corrected. The entire session lasted approximately 60 minutes per participant. The full procedure is outlined in Figure 3.

3.5 Data Analysis

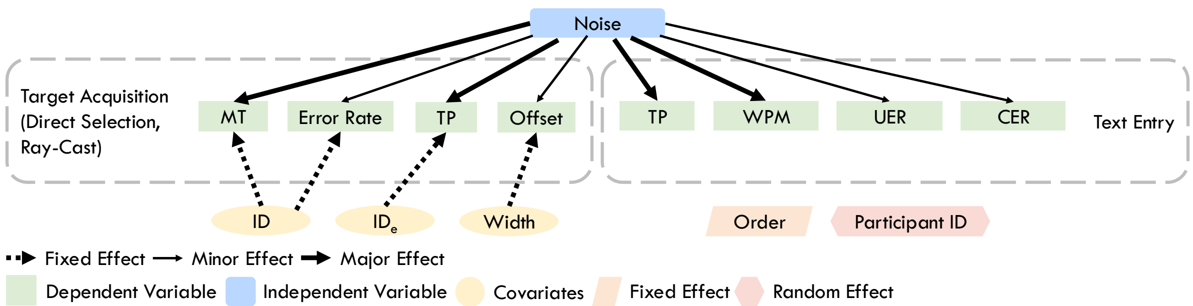


Fig. 4. Theorised causal directed acyclic graph. The thickness of the arrow represents the effect of the factor. The thicker the arrow, the stronger its effect.

We argue that ambient noise affect movement time, pointing offset, error rate, and text entry throughput at different levels as shown in Figure 4. To enhance the precision of our estimates, we include covariates such as ID, ID_e, and Width, which are known to impact movement time, pointing offset, and error rate. The order of conditions (Order) is incorporated into the models as a random effect to account for individual differences in fatigue and learning effects. Given our within-subject design with randomized condition orders, these factors do not confound the results [3, 21]. Participant ID is also modeled as a random effect to represent the hierarchical structure of the data. The variables included in the model are outlined below:

- **Noise:** A categorical variable indicating the noise participant experienced.
- **ID:** A numeric variable indicating the index of difficulty. This covariate was used to analyse movement time and error rate [20].
- **ID$_e$:** A numeric variable indicating the effective index of difficulty. This covariate was used to analyse target acquisition throughput [42].
- **Width:** A numeric variable indicating the width/diameter of the target. This covariate was used to analyse pointing offset [41].
- **Order:** A numeric variable indicating the order of the condition, which was treated as a fixed effect.
- **Participant ID:** A random effect used to model individual differences.

We use Python `3.12.0` and R `4.3.1` to clean and analyse the data. We used `Pandas 2.1.2`, `Numpy 1.26.1` (Python), and `brms 2.21.0`, `ggplot2 3.4.4` (R) libraries for our data analysis. We semantically analysed the video recordings from each participant and summarised their interview answers.

We employed Bayesian statistical methods in our analysis due to their enhanced flexibility, capacity to quantify uncertainty, and ability to facilitate future work to build upon it [8, 47]. This method is widely used in HCI research [8, 65]. We followed standard Bayesian analysis approach to conduct our analysis [15, 28, 74]. We fit our models using the `brms` package [9], which implements Bayesian multilevel models in R using the Stan probabilistic programming language [11]. We used regularising priors designed to be sceptical of implausibly large effect sizes [36]. We assessed the convergence and stability of the Markov Chain Monte Carlo sampling with

R-hat, which should be lower than 1.01 [75] and the Effective Sample Size (ESS), which should be greater than 1000 [9]. All of our estimates fit these criteria. We report the posterior means of parameter estimates, the error of these estimates, and the upper and lower bounds of the 95% compatibility interval (i.e., credible interval, CI) [9]. This compatibility interval indicates the range of values where there is a 95% probability that the true value falls within. For full transparency, all our analysis scripts and results can be found in the supplementary material.

We treat all hypotheses as "in principle positive" before seeing the data, so that the posterior probability directly quantifies our updated belief in a positive effect after observing the data [12]. This framing follows standard Bayesian formulations and ensures consistency: the posterior probability directly maps onto our decision criterion without invoking p-values or binary "significance" labels [44]. We report the hypothesis test results using posterior probability, which reflects the updated belief about a hypothesis of the positive effect after seeing the data, considering both the prior and the likelihood [30]. We interpret the values following the approach by Li et al. [36], where the approach was proved by Kruschke [27], considering values higher than 90% as accepting the hypothesis; 60% - 90% as indicating some evidence, but it may not be strong enough; 40% - 60% as suggesting the hypothesis is unlikely; 10% - 40% as indicating some evidence about negative effect, but it may not be strong enough; and below 10% as indicating strong evidence about negative effect. We note that p-values are not used in Bayesian statistics, and no claims about "statistical significance" should be derived from our results.

We use thematic analysis to process and analyse interview materials. Two authors repeatedly reviewed the data. Initial coding and theme generation were conducted by one author, and subsequently reviewed and refined by both authors. The final analysis resulted in the identification of the following three themes in Section 4.4.

4 RESULTS

We begin by presenting the overarching results in Section 4.1, where we summarise the mean performance and overall effects across all conditions. Thereafter, Section 4.2 and Section 4.3 provide the detailed breakdown model estimates (on the log scale), uncertainty measures (credible intervals and posterior probabilities), and text interpretation of each metric, which offer a detailed, metric-specific interpretation of each finding. All the model results tables were provided in Appendix C and supplementary materials to reduce redundancy and improve readability. Table 1, Table 2, and Table 3 serve as the primary reference tables. We further highlighted increasing trends in yellow, decreasing trends in blue, and used black to indicate no effect.

4.1 General Findings

It is important to note that Table 1 reflects the modelled estimates accounting for interaction effects between conditions and participant-level covariates. As such, the "pure" effect of each condition in isolation is not always directly visible in the raw tables, since the interaction terms modulate how conditions influence performance across different participant characteristics. For this reason, Table 1 is presented as the main reference table, as it summarises the overall effects of conditions after accounting for these interactions. The detailed tables 5-11 in the Appendix C report the full model outputs and uncertainty measures, and should be consulted alongside the summary Table 1 for a complete interpretation.

Table 1 provides a summary of the estimated mean values from the model. In the direct-selection task, baseline movement time was 597 ms, rising to 612 ms in outdoor urban noise condition (2.51% increase compared to silence condition), while meaningless speech yielded the fastest responses (589 ms, 1.34% decrease compared to silence condition). Throughput increased under fast-tempo music (1.50 bit/s, 2.74% increase compared to silence condition) and decreased most in outdoor noise (1.42 bit/s, 2.74% decrease compared to silence condition). For ray-casting, movement times were longest with fast music (1388 ms, 4.44% increase compared to silence condition) and meaningful speech (1369 ms, 3.01% increase compared to silence condition). In text entry, uncorrected and

Table 1. The estimated mean values and the corresponding trends of each metric per condition. MeaningfulS refers to meaningful speech; MeaninglessS refers to meaningless speech; MT refers to movement time in milliseconds; Offset refers to pointing offset in millimetres; Error refers to error rate in percentage; TP refers to throughput in bits/s; UER refers to uncorrected error rate in percentage; CER refers to corrected error rate in percentage; CPM refers to characters per minute.

	Direct Selection				**Ray-Casting**				**Text Entry**			
Condition	**MT**	**Offset**	**Error**	**TP**	**MT**	**Offset**	**Error**	**TP**	**TP**	**UER**	**CER**	**CPM**
Silence	597	22.10	0.44	1.46	1329	38.85	1.26	0.87	5.41	2.96	3.47	15.97
MeaningfulS	601–	22.23–	0.39–	1.45↓	1369↑	38.63↓	0.95–	0.85↓	5.47↑	2.97–	3.64↑	16.32↑
MeaninglessS	589↓	21.86↓	0.52–	1.47↑	1333–	38.27↓	0.88–	0.87–	5.12↓	3.93↑	3.38↓	15.54↓
Fast Music	600–	21.37↓	0.38–	1.50↑	1388↑	38.06↓	0.97–	0.84↓	5.37↓	3.60↑	3.57↑	15.84↓
Slow Music	591↓	21.20↓	0.14–	1.46–	1319↓	38.94–	1.06–	0.88↑	5.16↓	3.82↑	3.71↑	15.54↓
Urban Indoor	595↓	21.40↓	0.17–	1.44↓	1377↑	37.53↓	0.69–	0.82↓	5.59↑	2.94–	3.53–	16.67↑
Urban Outdoor	612↑	20.49↓	0.18–	1.42↓	1337–	38.26↓	0.88–	0.87–	5.41–	4.97↑	3.58↑	16.28↑

corrected error rates were lowest indoors (2.94% UER, 3.53% CER) and highest outdoors (4.97% UER), with increased throughput (5.59 bit/s) and speed (16.67 CPM) under indoor urban noise.

4.2 Target Acquisition

In the target acquisition task, we collected 24,704 records, which encompasses 11 targets per task per person across 7 conditions (silence, fast tempo music, slow tempo music, indoor ambient noise, outdoor ambient noise, meaningful speech, and meaningless speech), 6 IDs, 2 task types (direct selection and ray-casting), and XX participants, along with instances of incorrect selections.

4.2.1 Movement Time. For the analysis of movement time, we retained only the trials where targets were correctly selected; incorrect selections were reserved for error rate evaluation [36]. The movement time data follows a shifted Log-normal distribution [53, 55], which aligned with previous studies about analysing encumbrance and walking situational impairments [36]. We applied Bayesian regression on the log-transformed movement times to assess the effects of ambient noise.

Table 5 reports the posterior means of parameter estimates, the errors of these estimates, and the upper and lower bounds of the 95% compatibility/credible interval of movement time. We visualised the mean values and the 95% CI for the movement time per condition in Figure 5 in Appendix B. The results indicate that meaningless speech condition has a substantial negative effect on the movement time, where the posterior probability is 12%. Similarly, the urban outdoor condition has a substantial positive effect on the movement time, where the posterior probabilities is 95%. The meaningful and fast music conditions do not have substantial impact on the movement time, where the posterior probabilities are 43% and 46% respectively. Moving to the target acquisition via ray-casting task, the posterior probabilities indicate that meaningful speech, fast music, and urban indoor conditions have a positive impact on the movement time.

4.2.2 Pointing Offset. Similar to the movement time, we filtered our data and kept only the data with correctly selected targets; uncorrected selections were used to analyse the error rate [36]. The pointing offset follows a shifted Log-normal distribution [53, 55], which aligned with previous studies about analysing encumbrance and walking situational impairments [36]. We used Bayesian regression to model the effect of ambient noise on pointing offset in log scale.

Table 6 reports the posterior means of parameter estimates, the errors of these estimates, and the upper and lower bounds of the 95% CI of the pointing offset. We visualised the mean values and the 95% CI for the pointing offset per condition in Figure 6 in Appendix B. In the target acquisition via direct selection task, the model suggests that the meaningful speech condition has a substantial negative effect on pointing offset (-0.17, SD = 0.09, CI = [-0.35, 0.00]), where the posterior probability is 3%. The fast music and urban outdoor conditions also have a substantial negative effect on pointing offset via direct selection as indicated from the posterior probabilities (5% and 9% respectively).

4.2.3 Error Rate. We use Bayesian regression with Poisson distribution to model the effect of ambient noise on the error rate. Table 7 reports the posterior means of parameter estimates, the errors of these estimates, and the upper and lower bounds of the 95% CI of the error rate. We visualised the mean values and the 95% CI for the error rate per condition in Figure 7 in Appendix B. Fitts' law assumes a 4% error rate according to its information theory basis [40, 69]. Thus, since the values under all conditions are below 4%, there is no substantial effect on error rate across all conditions. In target acquisition via direct selection task, the posterior probabilities indicate that slow music and urban outdoor conditions have a substantial negative effect on error rate (4% for both), which means these conditions reduce error rate.

4.2.4 Target Acquisition Throughput. To analyse target acquisition throughput, we kept all data from target acquisition tasks as the metric can evaluate both speed and accuracy. The throughput follows a shifted Log-normal distribution [53, 55]. We used Bayesian regression to model the effect of ambient noise on target acquisition throughput in a log scale.

Table 8 reports the posterior means of parameter estimates, the errors of these estimates, and the upper and lower bounds of the 95% CI. We visualised the mean values and the 95% CI for the target acquisition throughput per condition in Figure 8 in Appendix B. In the target acquisition via direct selection task, the posterior probabilities suggest that meaningful speech (posterior probability = 0.01%), urban indoor (posterior probability = 0.04%), and urban outdoor (posterior probability = 0%) indicate negative effects on target acquisition throughput, which lead to a decrease in throughput. Similarly, the fast music condition indicate negative effects on target acquisition throughput in the target acquisition via ray-casting task, which lead to a decrease in throughput (posterior probability = 0.9%). Similar to the direct selection task, in the ray-casting task, the urban indoor condition indicate negative effects on throughput, which lead to a decrease in throughput (posterior probability = 1.3%).

4.3 Text Entry

Each participant completed 35 sentences during the study (5 sentences per condition), resulting in a total of 1,050 sentences collected during the experiment. During data cleaning, we identified 7 sentences that were left empty due to participants inadvertently pressing the enter key and moving on to the next sentence. Following previous study's method [36], these empty entries, representing 0.67% of the dataset, were removed as they do not provide a measure of performance. User performance in the text entry task is evaluated based on throughput.

4.3.1 Text Entry Throughput. We utilized Bayesian regression with a Gaussian distribution to investigate the effects of ambient noise on typing throughput. This model quantifies throughput as the number of characters typed per minute while incorporating the error rate on a normal scale.

Table 9 reports the posterior means of parameter estimates, the errors of these estimates, and the upper and lower bounds of the 95% CI. We visualised the mean values and the 95% CI for the text entry throughput per condition in Figure 9 in Appendix B. The model suggests that meaningless condition has a negative effect on text entry throughput (-0.30, SD = 0.14, CI = [-0.57, -0.02]). The posterior probability shows a 2% probability leading to higher throughput, which leads to a decrease in throughput. Similarly, the slow music condition shows a 4%

probability leading to higher throughput, which leads to a decrease in throughput. The urban indoor condition has a 89% probability leading to higher throughput.

4.3.2 CPM. We further analyses the text entry task using characters per minute (CPM). The CPM follows a shifted Log-normal distribution. We used Bayesian regression to model the effect of ambient noise on CPM.

Table 11 reports the posterior means of parameter estimates, the errors of these estimates, and the upper and lower bounds of the 95% CI. We visualised the mean values and the 95% CI for the characters per minute per condition in Figure 9a in Appendix B. The posterior probability indicates that urban indoor condition has a strong positive effect on CPM (94%). Besides, the meaningless and slow music condition has a negative effect on CPM (17%). Other conditions show a moderate or no effect on CPM.

4.3.3 UER and CER. We further evaluate uncorrected error rate (UER) and corrected error rate (CER) for text entry task. Both error rates are characterized by a zero-inflated Beta distribution. Bayesian regression was applied to examine the effects of ambient noise on UER and CER.

Table 10 reports the posterior means of parameter estimates, the errors of these estimates, and the upper and lower bounds of the 95% CI. We visualised the mean values and the 95% CI for the UER and CER per condition in Figure 10 in Appendix B. The posterior probabilities show a strong positive effect on UER under meaningless, slow music, and urban outdoor condition (89%, 85%, and 98% respectively). Similarly, the slow music condition shows a positive effect on CER as well (78%). Other conditions have only a moderate or no effect on UER and CER.

4.4 Summary of Interview

4.4.1 Subjective Impact of Ambient Sound on Task Performance. Multiple participants reported strong aversions to meaningful speech and urban ambient outdoor noise. Compared to meaningless speech, most participants indicated that meaningful speech was more distracting and cognitively demanding, as they could comprehend the spoken content and, in some cases, found themselves drawn to listen. For example, participants P20, P22, and P23 commented that hearing speech in their native language made it difficult to focus on the task: “Because I understand what is being said, I get distracted.” Similarly, P24 described frequent attentional shifts toward the audio when exposed to meaningful speech. This suggests that when linguistic information is semantically accessible, it competes with cognitive resources necessary for task execution.

Urban outdoor sounds were also described as highly disruptive. As P30 and P2 stated, “Outdoor urban noise makes me feel irritated,” while P23 shared, “It makes me imagine the environment I am in.” These remarks highlight that multi-source, chaotic soundscapes—such as traffic and crowded public spaces—can interfere with sensory processing and potentially induce psychological discomfort or emotional resistance.

Participants’ responses suggest that both the semantic load of the sound and their familiarity with the auditory context significantly shape their experience. These findings underscore the need for MR system designers to consider the interaction between semantic salience and attentional guidance when integrating auditory elements into immersive environments.

4.4.2 The Regulatory and Supportive Role of Music. In contrast to speech and environmental noise, background music—both fast and slow tempo—was frequently associated with positive emotional regulation. In particular, slow-tempo music was repeatedly described as soothing and enjoyable: “Slow tempo music makes me feel the most comfortable, relaxed, and I enjoy it” (P11, P22, P24, P28). P5 specifically noted that such music helped induce a calm and centered state. These observations align with existing studies on music-induced meditative states, suggesting that slow-tempo music may enhance task performance by reducing heart rate variability and supporting sustained attention.

4.4.3 Individual Differences and Sound Adaptation Patterns. Participants' responses also revealed marked individual differences in their preferences, familiarity, and habitual exposure to various types of sounds. These divergences indicate that the impact of auditory stimuli on user performance is deeply intertwined with personal background factors such as linguistic ability, disciplinary training, and everyday listening behaviors. For example, P2 mentioned, "I usually listen to English while typing, so I'm used to speech." P14 shared that she lives next to a busy road and has become accustomed to such environments, suggesting that outdoor noise does not disrupt her; rather, she "works better" under such conditions. P30 remarked that indoor noise reminded her of a library, which made her feel more relaxed. These reflections suggest that the effects of ambient sound on perception and performance are not universal, but rather mediated by prior experiences, environmental habituation, and contextual associations. Consequently, the implementation of auditory stimuli in MR systems should consider personalized adaptability to accommodate a wider range of user sensitivities and behavioral tendencies.

5 DISCUSSION

Across noise types, structured and rhythmic sounds (speech, music) generally encourage speed–accuracy tradeoffs, whereas urban environments force users to slow down and heighten precision at the cost of throughput and comfort. These patterns underscore the need for MR interfaces to adapt dynamically to the listener's auditory context. Although the 95% credible interval does not show substantial effect from the results, the posterior probability of a effect can still be high because these two summaries capture different aspects of the posterior distribution. Even if a 95% credible interval for a parameter just barely includes zero, the bulk of the posterior density can lie well away from zero. In other words, the credible interval is a central-mass summary, not a direct statement about how much probability sits on one side of zero versus the other. Thus, a CI that just grazes zero can coexist with a high posterior probability, reflecting that only a small tail of the distribution extends into negative values while most of the mass—and hence our belief—lies firmly on the positive side and vice versa. In this section, we discuss the detailed effects of six ambient noises, elaborate on potential reasons behind the observed phenomenon, and contrast the findings with previous evaluated situational impairments on both MR and mobile interaction.

5.1 Effects of Ambient Noise on MR Interaction

Table 2 and Table 3 summarize the effects of noise observed in the target acquisition tasks and text entry task respectively, contrast them with other situational impairments and the effects of noise on mobile interaction.

5.1.1 Speech Condition. Across both target selection and text-entry tasks, ambient noise's impact depended heavily on its semantic content. In the target-acquisition study, meaningless speech improved direct-selection performance—reducing movement time (1.34% decrease), tightening offsets (1.09% decrease), and worsen throughput (0.68% increase). In text entry, meaningful speech increased throughput (1.11%), CER (4.90%), and CPM (2.19%), while meaningless decreased throughput (5.36%), CER (2.59%), and CPM (2.69%). Our quantitative results were reinforced by qualitative insights gathered from semi-structured interviews. For example, P20, P22, and P23 commented that hearing speech in their native language made it difficult to focus on the task. Interpreted through the lens of the Yerkes–Dodson law [80], these results suggest that semantically rich noise can heighten arousal enough to optimise highly automated motor routines (such as direct-selection pointing), but when combined with meaningful linguistic content it overloads the phonological loop, forcing a speed–accuracy trade-off that particularly undermines language production and more cognitively demanding selection methods [14, 19]. The practical implication for interface design is two-fold. First, in environments where rapid, routine interactions dominate (e.g., select menus, control-room dashboards), introducing low-semantic babble may actually benefit user throughput. Second, in contexts requiring high-precision text entry or complex selection (e.g., CAD tools),

Table 2. Percentage change in target acquisition tasks' metrics relative to silent condition.

	Target Acquisition (Direct Selection)				Target Acquisition (Ray-Casting)			
Condition	**MT**	**Offset**	**Error**	**TP**	**MT**	**Offset**	**Error**	**TP**
Ambient Noise on MR Interaction								
Meaningful	–	–	–	↓ 0.68%	↑ 3.01%	↓ 0.57%	–	↓ 2.30%
Meaningless	↓ 1.34%	↓ 1.09%	–	↑ 0.68%	–	↓ 1.49%	–	–
Fast Music	–	↓ 3.30%	–	↑ 2.74%	↑ 4.44%	↓ 2.03%	–	↓ 3.45%
Slow Music	↓ 1.01%	↓ 4.07%	–	–	↓ 0.75%	–	–	↑ 1.15%
Urban Indoor	↓ 0.34%	↓ 3.17%	–	↓ 1.37%	↑ 3.61%	↓ 3.40%	–	↓ 5.75%
Urban Outdoor	↑ 2.51%	↓ 7.29%	–	↓ 2.74%	–	↓ 1.52%	–	–
Encumbrance on MR Interaction								
1.0 kg	↑ 28%	–	–	↓ 22%	↑ 27%	–	–	↓ 20%
Walking on MR Interaction								
Constant Speed	↑ 21%	↑ 16%	–	↓ 16%	↑ 63%	↑ 17%	↑ 8.4%	↓ 32%
Ambient Noise on Mobile Interaction								
Meaningful	–	–	✗	✗	✗	✗	✗	✗
Meaningless	–	–	✗	✗	✗	✗	✗	✗
Fast Music	↓	↑	✗	✗	✗	✗	✗	✗
Slow Music	↓	↑	✗	✗	✗	✗	✗	✗
Urban Indoor	↓	–	✗	✗	✗	✗	✗	✗
Urban Outdoor	↓	–	✗	✗	✗	✗	✗	✗

↑ increased relative to baseline
↓ decreased relative to baseline
- no substantial effect was observed
✗ not applicable

minimising intelligible background speech, or providing active noise cancellation targeted at speech frequencies, could substantially reduce error rates and cognitive load.

5.1.2 Music Condition. With fast-tempo music, the offset improved by 3.30% and throughput increased 2.74% under direct selection task. However, in ray-casting task, the movement time increased 4.44% and throughput decreased 3.45%. It further decreased text entry throughput (0.74%) and CPM (0.81%) while increase UER (21.62%) and CER (2.88%). Slow music produced the largest accuracy gain in direct selection offsets improved by 4.07% and movement time decreased by 1.01%, while throughput stayed roughly the same; in ray-casting it added a small 1.15% boost in throughput despite a 0.75% slowdown. But for text entry slow music decreased performance most: throughput fell 4.62%, uncorrected errors increased 29.05%, corrected errors increased 6.92%, and CPM decreased 2.69%.

This could be explained using Dynamic Attending Theory [31], where rhythmic stimuli serve to entrain our internal attentional pulses, effectively syncing moments of heightened sensitivity to the temporal structure of incoming events [16]. Under a steady, slow beat, these pulses become more predictable, sharpening our ability to time and guide reaching movements and thereby improving both pointing accuracy and speed. In contrast, a rapid beat can throw this synchronization off balance, desynchronizing attentional peaks and valleys and making it harder to coordinate the precise timing required for tasks like clicking small targets or typing—ultimately

Table 3. Percentage change in text entry task's metrics relative to silent condition.

Condition	TP	UER	CER	CPM
Ambient Noise on MR Interaction				
Meaningful	↑ 1.11%	–	↑ 4.90%	↑ 2.19%
Meaningless	↓ 5.36%	↑ 32.77%	↓ 2.59%	↓ 2.69%
Fast Music	↓ 0.74%	↑ 21.62%	↑ 2.88%	↓ 0.81%
Slow Music	↓ 4.62%	↑ 29.05%	↑ 6.92%	↓ 2.69%
Urban Indoor	↑ 3.33%	–	–	↑ 4.38%
Urban Outdoor	–	↑ 67.91%	↑ 3.17%	↑ 1.94%
Encumbrance on MR Interaction				
1.0 kg	↓ 17%	↑ 50%	–	↓ 16%
Walking on MR Interaction				
Constant Speed	↓ 51%	↑ 28%	↑ 68%	↓ 45%
Ambient Noise on Mobile Interaction				
Meaningful	✗	–	✗	↑
Meaningless	✗	–	✗	↑
Fast Music	✗	–	✗	–
Slow Music	✗	–	✗	–
Urban Indoor	✗	–	✗	↑
Urban Outdoor	✗	–	✗	↑

↑ increased relative to baseline
↓ decreased relative to baseline
- no substantial effect was observed
✗ not applicable

increasing error rates. Practically, interface designers might leverage slow-tempo background tracks in environments dominated by routine pointing gestures—such as kiosks or simple menu navigation—to enhance accuracy without slowing users down. Conversely, when tasks require multitouch gestures, drag-and-drop, or extensive text input, minimising background music or allowing users to select tracks with tempos matched to their task demands could reduce errors and cognitive strain.

5.1.3 Urban Noise Condition. Under urban-indoor noise condition, participants showed improvement in precision, where direct-selection movement time decreased 0.34% and offset decreased 3.17%. Typing throughput increased 3.33% and CPM increased 4.38%). By contrast, the unpredictable racket of urban-outdoor settings slowed direct selection movement time by 2.51%, decreased offset by 7.29%, and decreased throughput by 2.74%). Our quantitative results were reinforced by qualitative insights, where urban outdoor sounds were described as highly disruptive (P30 and P2).

Several studies show that indoor urban ambient noise can boost performance on both creative and routine cognitive tasks. For example, Mehta et al. [48] demonstrated that listening to a recorded coffee shop soundtrack at about 70 dB enhanced participants' creative task performance by roughly 25%, compared with quieter (50 dB) or louder (85 dB) conditions. Similarly, a recent experimental study of neurotypical young adults found that introducing white noise at 45 dB led to 5-10% gains in sustained attention, accuracy, and task speed [2]. In real-world settings, users feel more focused when working in cafe's or open-plan co-working spaces, and some companies now offer “coffee-shop” soundtracks in quiet zones to recreate these performance benefits without leaving the office. In practical terms, these insights point to the potential benefits of integrating “indoor urban” soundtracks into settings where precision pointing or text entry is critical, thereby harnessing mild auditory stimulation to support fluency. Conversely, they highlight the importance of mitigating uncontrolled outdoor noise in workspaces to preserve the focused, uninterrupted attention necessary for both rapid motor routines and sustained typing tasks.

5.1.4 Comparisons between Different Ambient Noises. Across all six ambient conditions, we see a clear trade-off between simple motor tasks and more complex control or cognitive tasks. Meaningful speech left direct pointing mostly unchanged but slowed and reduced throughput in ray-casting, while meaningless speech actually sharpened direct-selection speed and accuracy without helping ray-casting. Fast music tightened direct-selection offsets and boosted throughput but made ray-casting slower and less efficient; slow music steadied direct pointing yet only slightly helped ray-casting. The urban-indoor noise produced small gains across both pointing and typing, whereas the unpredictable urban-outdoor settings consistently decrease pointing performance and drove uncorrected typing errors through the roof.

5.2 Contrasting Situational Impairments on MR and Mobile Interaction

5.2.1 Contrasting with the Effects of Encumbrance and Walking on MR Interaction. To advance our understanding of how situational impairments affect user performance in MR, we compare the effects of ambient noise to the effects of movement and physical encumbrance presented in [36]. This approach enables a more comprehensive evaluation of MR usability under real-world conditions. By incorporating findings from studies with comparable methodologies, we establish a consistent framework for comparison across different types of situational impairments. Consistent with established protocols [36, 60, 61, 63], we quantify the effects of situational impairments as percentage changes relative to corresponding baseline conditions. This normalisation facilitates fair cross-study comparisons by allowing us to directly compare the magnitude of different situational impairments across similar tasks and conditions.

Although ambient noise produced measurable changes in movement time, accuracy and throughput across all tasks (Table 2 and Table 3), these effects were modest in magnitude compared to those induced by encumbrance and walking [36]. A 1 kg weight on the hand increased direct-selection movement times by roughly 28%, and ray-casting by 27%, whereas meaningful or urban noise typically shifted movement times by a small arrow or left them unchanged [36]. Pointing accuracy under encumbrance also suffered—throughput fell by 22% for direct-selection and 20% for ray-casting [36]. By contrast, meaningful or urban noise typically shifted movement times by only a few percentage points—or not at all—because auditory distractions chiefly tax cognitive resources rather than gross motor dynamics.

Locomotion proved even more disruptive: walking slowed direct-selection by 21% and ray-casting by 63%, cut direct-selection throughput by 16% and ray-casting by 32%, and reduced typing throughput by over 50% [36]. By comparison, fast-tempo or indoor urban noise rarely altered throughput by more than 10%.

When encumbrance and walking were combined, movement times more than doubled for ray-casting and typing throughput dropped by nearly 60%, dwarfing any single noise effect. In sum, while ambient sounds do influence MR interaction—often through subtle speed-accuracy trade-offs—their overall impact is small relative to the physical constraints imposed by added weight or the demands of locomotion.

5.2.2 Contrasting with the Effects of Ambient Noise on Mobile Interaction. In MR, hearing meaningful speech had only subtle effects on direct-selection and ray-casting: users moved their hands just as quickly in direct-selection but pointed more precisely (large drop in offset) without increasing errors, though overall information throughput fell sharply. When typing in MR, participants actually sped up (higher throughput and CPM) but made more character-level errors. In contrast, on smartphones, meaningful speech neither slowed down taps nor degraded targeting or icon recall, but it did slow text entry: time per character rose significantly, even though tapping speed, icon search accuracy and typing-error rates remained unchanged. MR text entry under meaningless speech was markedly slower. Similarly, on smartphone interaction, meaningless speech did slow typing: time per character increased noticeably, with no change in overall typing-error rates.

When fast-tempo music played in MR, direct-selection movement times jumped but spatial accuracy improved, producing a modest boost in throughput. Ray-casting saw no timing change but suffered large losses in offset and throughput despite a small error-rate reduction. Text entry slowed slightly and error rates increased at both the uncorrected and corrected levels. On smartphones, fast music increased circle-tapping times significantly at the expense of larger offsets, while it did not affect icon recall or search errors, and lengthened the time per character entry, again with no increase in typing errors. Slow music in MR left direct-selection speed unchanged but yielded strong improvements in pointing offset and error rate at the cost of a large throughput decline. Ray-casting became slightly faster and more precise but also generated more errors and lower throughput. Text entry suffered severe slowdowns and elevated error rates. On mobile, slow music accelerated tapping times significantly but introduced greater offsets, and slowed text entry (higher time per character), with typing-error rates unaffected.

In MR, indoor urban noise strongly slowed direct-selection, degraded accuracy and error rate, and caused a pronounced throughput drop. Ray-casting movements also slowed moderately, but error rates fell sharply even as throughput declined. Text entry, surprisingly, became faster (both throughput and characters per second rose) without any change in error rates. On smartphones, indoor noise hastened circle taps, and significantly slowed text entry (higher time per character), while tapping accuracy and typing-error rates remained stable. Under outdoor urban noise in MR, users moved much more slowly in direct-selection but achieved much better precision; ray-casting timing and throughput were unaffected. MR text entry showed faster typing speed but a sharp rise in both uncorrected and corrected errors. On mobile devices, outdoor noise significantly sped up tapping times, did not affect offsets or icon recall, and slowed text entry (higher time per character), again without changing typing-error rates.

5.3 Addressing the Effects of Ambient Noise on MR Interaction

Our study reveals that ambient noise can influence MR interactions in certain conditions. Prior research has indicated that identifying environmental challenges is essential for the successful adaptation of user interfaces [62, 63]. If future MR devices would achieve widespread adoption for daily use, addressing these challenges is a crucial step [36]. Tigwell et al. [73] proposed two steps to address situational impairments: sensing and adapting.

5.3.1 Sensing Ambient Noise. Recognizing these situational impairments automatically during devices use enables interfaces to adjust appropriately, making device interaction more context-sensitive [63]. Solutions should leverage the headset's integrated sensors rather than depend on external devices or instruments that could impede users' interaction with the interface [36]. In line with this idea, Sarsenbayeva et al. [63] proposed that the mobile device's integrated microphone can serve to detect noise-related impairments. Additionally, a classifier could be implemented to differentiate between various noise types [63]. These can be further adapted to MR devices to better sense the noise environment.

5.3.2 Adapting Ambient Noise. From a hardware perspective, incorporating noise cancellation functionality could potentially improve MR device design to mitigate ambient noise. For example, a noise-cancelling earmuff could be integrated to the side of the headset so that it can be worn when needed to block out external noise. However, it is important to note that complete noise suppression may compromise situational awareness, especially during locomotion; hence, devices should support transparency or ambient-pass-through modes to preserve critical environmental cues for user safety.

From a software perspective, MR interfaces could adapt based on the identified noise condition. For instance, if urban noise or speech—both known to impair text entry on mobile interaction [63]—is detected while a user is typing, the device might switch to an alternative interface designed to counteract the adverse effects of these noise-induced impairments, such as the "WalkType" interface proposed by Goel et al. [22] or the "Fat thumb" technique proposed by Boring et al. [7]. Moreover, by detecting the user's activity (e.g., walking vs. standing still) and the noise profile, the system could dynamically toggle between full ambient noise cancellation and transparency modes to optimise both performance and safety. However, there is still limited research that explores ways to adapt MR interfaces based on the impact of different noise types on user performance.. Future studies should therefore explore not only novel input methods tailored to different noise conditions, but also adaptive audio strategies that balance noise suppression with the preservation of essential ambient sounds to ensure user safety in diverse contexts.

5.4 Limitations and Future Work

We acknowledge several limitations in our study. First, the study settings were strictly controlled. The controlled setting may not fully capture the diverse and simultaneous noise types encountered in a natural environment. For example, urban indoor noise might include fast or slow music in a shopping mall. Besides, we did not include music with lyrics, which combines the effects of both musical composition and speech. Nevertheless, we argue it was necessary to control for these variables in order to eliminate any potential confounding effects and to systematically compare their specific effects on performance [36, 60]. Since we know the prominent effects of some types of ambient noise, we hypothesise that the interaction effects will be more pronounced in real-world scenarios.

Additionally, our study focused solely on target acquisition and text entry tasks, even though we recognize that real-world MR applications often involve more complex activities. We controlled the task type to isolate the effects of ambient noise on basic MR functions. Given the impacts observed on these simple tasks, we expect these effects to be even more pronounced with more complex MR tasks. Future work should broaden the scope of current tasks and place greater emphasis on real-world scenarios. We also restricted our participants to use only the index finger of their dominant hand to perform the tasks. This restriction was necessary to ensure a fair comparison across different conditions [36, 63]. Moreover, by restricting the interaction technique we created a more comparable setting between participants.

A further limitation concerns our participant pool. Our sample was restricted to individuals with a bachelor's degree or higher, within a relatively narrow age range. While our initial analysis suggested that education and age did not exert significant effects on performance, the absence of data from a more diverse participant group makes it difficult to claim with certainty that these variables have no influence. We relied on participant random effects to account for unmeasured individual differences, but future studies should include broader age, educational, and cultural backgrounds to better establish the generalisability of our findings.

Finally, we note a limitation regarding the comparison of our findings with those of prior situational impairment studies. While we argue that it is more insightful to examine how different situational impairments impact MR performance [60], differences in study design, task implementation, and participant samples may introduce uncontrolled factors that limit the validity of direct comparisons. Although we normalized our results to a baseline condition [60, 61, 63], such normalization may not fully account for non-proportional effects across impairments. Thus, any cross-study comparisons should be interpreted cautiously. Future work could aim to establish unified experimental protocols to better support systematic comparisons across different situational impairments.

6 CONCLUSION

In this study, we examine the effects of six types of ambient noise on MR interaction performance in target acquisition (direct selection and ray-casting) and text entry tasks. Our results indicated that meaningless speech decreased text entry throughput by 5.36%, fast-tempo music increased movement time by 4.44%, slow-tempo music increased offset by 4.07%, urban indoor noise increased typing throughput by 3.33%, and urban outdoor noise decreased throughput by 2.74%. We found that the target acquisition throughput decreased substantially under all ambient noise conditions except meaningless speech and fast-tempo music condition. Besides, meaningless speech, fast-tempo music, and slow-tempo music conditions decreased text entry throughput. Our findings further indicate that, when compared with encumbrance and walking conditions, the impact of ambient noise was relatively minor. We emphasize the importance of factoring situational impairments into MR interactions and present potential strategies for sensing and adapting to their effects. Overall, our results broaden our understanding of how ambient noise affects MR interactions and provide valuable contributions to the expanding research on SIIDs in the context of MR.

Acknowledgments

Dr. Withana is a recipient of an Australian Research Council Discovery Early Career Award (DECRA) - DE200100479 funded by the Australian Government. This work is supported by JST Moonshot R&D Program "Cybernetic being" Project (Grant number JPMJMS2013), and JST Doctoral Grant (Keio-SPRING, Grant number JPMJMS2123). We sincerely thank our participants for their valuable time and the reviewers for their constructive feedback.

References

[1] Amy FT Arnsten and Patricia S Goldman-Rakic. 1998. Noise stress impairs prefrontal cortical cognitive function in monkeys: evidence for a hyperdopaminergic mechanism. *Archives of general psychiatry* 55, 4 (1998), 362–368.

[2] Mohamad Awada, Burcin Becerik-Gerber, Gale Lucas, and Shawn Roll. 2022. Cognitive performance, creativity and stress levels of neurotypical young adults under different white noise levels. *Scientific Reports* 12, 1 (2022), 14566.

[3] R Harald Baayen, Douglas J Davidson, and Douglas M Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language* 59, 4 (2008), 390–412.

[4] Simon Banbury and Dianne C Berry. 1998. Disruption of office-related tasks by speech and office noise. *British journal of psychology* 89, 3 (1998), 499–517.

[5] Simon P Banbury, William J Macken, Sébastien Tremblay, and Dylan M Jones. 2001. Auditory distraction and short-term memory: Phenomena and practical implications. *Human factors* 43, 1 (2001), 12–29.

[6] Jens Blauert. 1997. *Spatial hearing: the psychophysics of human sound localization.* MIT press.

[7] Sebastian Boring, David Ledo, Xiang'Anthony' Chen, Nicolai Marquardt, Anthony Tang, and Saul Greenberg. 2012. The fat thumb: using the thumb's contact size for single-handed mobile interaction. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services.* 39–48.

[8] Joe Brailsford, Frank Vetere, and Eduardo Velloso. 2024. Exploring the Association between Moral Foundations and Judgements of AI Behaviour. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* 1–15.

[9] Paul-Christian Bürkner. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software* 80 (2017), 1–28.

[10] Kelly Caine. 2016. Local standards for sample size at CHI. In *Proceedings of the 2016 CHI conference on human factors in computing systems.* 981–992.

[11] Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus A Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. Stan: A probabilistic programming language. *Journal of statistical software* 76 (2017).

[12] George Casella and Roger L Berger. 1987. Reconciling Bayesian and frequentist evidence in the one-sided testing problem. *J. Amer. Statist. Assoc.* 82, 397 (1987), 106–111.

[13] Edwin John Crossman. 1957. *Factors involved in the predator-prey relationship of rainbow trout (Salmo gairdneri Richardson) and redside shiners (Richardsonius balteatus (Richardson)) in Paul Lake, British Columbia.* Ph. D. Dissertation. University of British Columbia.

[14] Daniel Deigård. 2012. The Effect of Acute Background Noise on Recognition Tasks.

[15] Sarah Depaoli and Rens Van de Schoot. 2017. Improving transparency and replication in Bayesian statistics: The WAMBS-Checklist. *Psychological methods* 22, 2 (2017), 240.

[16] Carolyn Drake, Mari Riess Jones, and Clarisse Baruch. 2000. The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. *Cognition* 77, 3 (2000), 251–288.

[17] Tafadzwa Joseph Dube and Ahmed Sabbir Arif. 2019. Text entry in virtual reality: A comprehensive review of the literature. In *Human-Computer Interaction. Recognition and Interaction Technologies: Thematic Area, HCI 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings, Part II 21*. Springer, 419–437.

[18] Tafadzwa Joseph Dube and Ahmed Sabbir Arif. 2020. Impact of key shape and dimension on text entry in virtual reality. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems*. 1–10.

[19] Richard D Edsell. 1976. Noise and social interaction as simultaneous stressors. *Perceptual and Motor Skills* 42, 3_suppl (1976), 1123–1129.

[20] Paul M Fitts. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology* 47, 6 (1954), 381.

[21] Andrew Gelman. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

[22] Mayank Goel, Leah Findlater, and Jacob Wobbrock. 2012. WalkType: using accelerometer data to accomodate situational impairments in mobile touch screen text entry. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2687–2696.

[23] Florian Heller, Davy Vanacken, Eva Geurts, and Kris Luyten. 2020. Impact of situational impairment on interaction with wearable displays. In *22nd international conference on human-computer interaction with mobile devices and services*. 1–5.

[24] Morris B Holbrook. 1981. Integrating compositional and decompositional analyses to represent the intervening role of perceptions in evaluative judgments. *Journal of Marketing Research* 18, 1 (1981), 13–28.

[25] Jin Huang and Byungjoo Lee. 2019. Modeling error rates in spatiotemporal moving target selection. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–6.

[26] Kari Kallinen. 2002. Reading news from a pocket computer in a distracting environment: effects of the tempo of background music. *Computers in Human Behavior* 18, 5 (2002), 537–551.

[27] John K Kruschke. 2013. Bayesian estimation supersedes the t test. *Journal of experimental psychology: General* 142, 2 (2013), 573.

[28] John K Kruschke. 2021. Bayesian analysis reporting guidelines. *Nature human behaviour* 5, 10 (2021), 1282–1291.

[29] Hannu Kukka, Jorge Goncalves, Kai Wang, Tommi Puolamaa, Julien Louis, Mounib Mazouzi, and Leire Roa Barco. 2016. Utilizing audio cues to raise awareness and entice interaction on public displays. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. 807–811.

[30] Ben Lambert. 2018. A student's guide to Bayesian statistics. (2018).

[31] Edward W Large and Mari Riess Jones. 1999. The dynamics of attending: How people track time-varying events. *Psychological review* 106, 1 (1999), 119.

[32] Grace Leslie, Asma Ghandeharioun, Diane Zhou, and Rosalind W Picard. 2019. Engineering music to slow breathing and invite relaxed physiology. In *2019 8th international conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 1–7.

[33] Tinghui Li, Pamuditha Somarathne, Zhanna Sarsenbayeva, and Anusha Withana. 2025. NeverLagging: Enhancing Virtual Reality Finger Tracking with a Physics-Inspired Time-Agnostic Graph Neural Network. In *The 37th Australian Conference on Human-Computer Interaction* (Sydney, Australia) *(OzCHI '25)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3764687.3769915

[34] Tinghui Li, Pamuditha Somarathne, Zhanna Sarsenbayeva, and Anusha Withana. 2025. TA-GNN: Physics Inspired Time-Agnostic Graph Neural Network for Finger Motion Prediction. *arXiv preprint arXiv:2503.13034* (2025).

[35] Tinghui Li, Brandon Syiem, Niels van Berkel, Juno Kim, Anusha Withana, and Zhanna Sarsenbayeva. 2025. Trends and Directions of Extended Reality Research in HCI: A Co-Word Analysis. In *The 37th Australian Conference on Human-Computer Interaction* (Sydney, Australia) *(OzCHI '25)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3764687.3769922

[36] Tinghui Li, Eduardo Velloso, Anusha Withana, and Zhanna Sarsenbayeva. 2025. Estimating the Effects of Encumbrance and Walking on Mixed Reality Interaction. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '25)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3706598.3713492

[37] Tinghui Li, Eduardo Velloso, Anusha Withana, and Zhanna Sarsenbayeva. 2025. Weight-Induced Consumed Endurance (WICE): A Model to Quantify Shoulder Fatigue with Weighted Objects. In *The 38th Annual ACM Symposium on User Interface Software and Technology* (Busan, Republic of Korea) *(UIST '25)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3746059.3747745

[38] Tinghui Li, Guangzheng Zhang, and Zhanna Sarsenbayeva. 2025. SOBO: An AI-Driven Small Object Boundary Detection Framework for Optimising VR Safety in Confined Spaces. In *The 37th Australian Conference on Human-Computer Interaction (OzCHI '25)*. ACM, Sydney, Australia. doi:10.1145/3764687.3769913

[39] Ruilin Liu, Tinghui Li, and Zhanna Sarsenbayeva. 2025. Understanding and Addressing Ambient Illumination as a Situational Impairment for Digital Devices. In *The 37th Australian Conference on Human-Computer Interaction (OzCHI '25)*. ACM, Sydney, Australia. doi:10.1145/3764687.3769905

[40] I Scott MacKenzie. 1992. Fitts' law as a research and design tool in human-computer interaction. *Human-computer interaction* 7, 1 (1992), 91–139.

[41] I Scott MacKenzie. 2018. Fitts' law. *The wiley handbook of human computer interaction* 1 (2018), 347–370.

[42] I Scott MacKenzie and Poika Isokoski. 2008. Fitts' throughput and the speed-accuracy tradeoff. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1633–1636.
[43] I Scott MacKenzie and R William Soukoreff. 2003. Phrase sets for evaluating text entry techniques. In *CHI'03 extended abstracts on Human factors in computing systems*. 754–755.
[44] Dominique Makowski, Mattan S Ben-Shachar, SH Annabel Chen, and Daniel Lüdecke. 2019. Indices of effect existence and significance in the Bayesian framework. *Frontiers in psychology* 10 (2019), 2767.
[45] Randi C Martin, Michael S Wogalter, and Janice G Forlano. 1988. Reading comprehension in the presence of unattended speech and music. *Journal of memory and language* 27, 4 (1988), 382–398.
[46] Heather McElrea and Lionel Standing. 1992. Fast music causes fast drinking. *Perceptual and Motor skills* 75, 2 (1992), 362–362.
[47] Richard McElreath. 2018. *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC.
[48] Ravi Mehta, Rui Zhu, and Amar Cheema. 2012. Is noise always bad? Exploring the effects of ambient noise on creative cognition. *Journal of Consumer Research* 39, 4 (2012), 784–799.
[49] Nils Meyer-Kahlen, Sebastian Schlecht, Sebastià V Amengual Garí, and Tapio Lokki. 2024. Testing Auditory Illusions in Augmented Reality: Plausibility, Transfer-Plausibility, and Authenticity. *Journal of the Audio Engineering Society* 72, 11 (2024), 797–812.
[50] Ronald E Milliman. 1982. Using background music to affect the behavior of supermarket shoppers. *Journal of marketing* 46, 3 (1982), 86–91.
[51] Alexander Ng, Stephen A Brewster, and John H Williamson. 2014. Investigating the effects of encumbrance on one-and two-handed interactions with mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1981–1990.
[52] Alexander Ng, John Williamson, and Stephen Brewster. 2015. The effects of encumbrance and mobility on touch-based gesture interactions for mobile phones. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 536–546.
[53] Karin Nieuwenhuizen and Jean-Bernard Martens. 2016. Advanced modeling of selection and steering data: beyond Fitts' law. *International Journal of Human-Computer Studies* 94 (2016), 35–52.
[54] Adrian C North, David J Hargreaves, and Sarah J Heath. 1998. Musical tempo and time perception in a gymnasium. *Psychology of Music* 26, 1 (1998), 78–88.
[55] Réjean Plamondon and Adel M Alimi. 1997. Speed/accuracy trade-offs in target-directed movements. *Behavioral and brain sciences* 20, 2 (1997), 279–303.
[56] Pamela Ravasio, Sissel Guttormsen Schär, and Helmut Krueger. 2004. In pursuit of desktop evolution: User problems and practices with modern desktop systems. *ACM Transactions on Computer-Human Interaction (TOCHI)* 11, 2 (2004), 156–180.
[57] Pierre Salamé and Alan Baddeley. 1989. Effects of background music on phonological short-term memory. *The Quarterly Journal of Experimental Psychology* 41, 1 (1989), 107–122.
[58] Zhanna Sarsenbayeva, Jorge Goncalves, Juan García, Simon Klakegg, Sirkka Rissanen, Hannu Rintamäki, Jari Hannu, and Vassilis Kostakos. 2016. Situational impairments to mobile interaction in cold environments. In *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing*. 85–96.
[59] Zhanna Sarsenbayeva, Vassilis Kostakos, and Jorge Goncalves. 2019. Situationally-Induced Impairments and Disabilities Research. *arXiv preprint arXiv:1904.06128* (2019).
[60] Zhanna Sarsenbayeva, Niels van Berkel, Danula Hettiachchi, Weiwei Jiang, Tilman Dingler, Eduardo Velloso, Vassilis Kostakos, and Jorge Goncalves. 2019. Measuring the effects of stress on mobile interaction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1 (2019), 1–18.
[61] Zhanna Sarsenbayeva, Niels Van Berkel, Weiwei Jiang, Danula Hettiachchi, Vassilis Kostakos, and Jorge Goncalves. 2019. Effect of ambient light on mobile interaction. In *Human-Computer Interaction–INTERACT 2019: 17th IFIP TC 13 International Conference, Paphos, Cyprus, September 2–6, 2019, Proceedings, Part III 17*. Springer, 465–475.
[62] Zhanna Sarsenbayeva, Niels van Berkel, Chu Luo, Vassilis Kostakos, and Jorge Goncalves. 2017. Challenges of situational impairments during interaction with mobile devices. In *Proceedings of the 29th australian conference on computer-human interaction*. 477–481.
[63] Zhanna Sarsenbayeva, Niels van Berkel, Eduardo Velloso, Vassilis Kostakos, and Jorge Goncalves. 2018. Effect of distinct ambient noise types on mobile interaction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (2018), 1–23.
[64] Zhanna Sarsenbayeva, Niels Van Berkel, Aku Visuri, Sirkka Rissanen, Hannu Rintamaki, Vassilis Kostakos, and Jorge Goncalves. 2017. Sensing cold-induced situational impairments in mobile interaction using battery temperature. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–9.
[65] Martin Schmettow. 2021. *New statistics for design researchers*. Springer.
[66] Andrew Sears, Min Lin, Julie Jacko, and Yan Xiao. 2003. When computers fade: Pervasive computing and situationally-induced impairments and disabilities. In *HCI international*, Vol. 2. 1298–1302.
[67] Claude Elwood Shannon. 1948. A mathematical theory of communication. *The Bell system technical journal* 27, 3 (1948), 379–423.
[68] R William Soukoreff and I Scott MacKenzie. 2003. Metrics for text entry research: An evaluation of MSD and KSPC, and a new unified error metric. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 113–120.

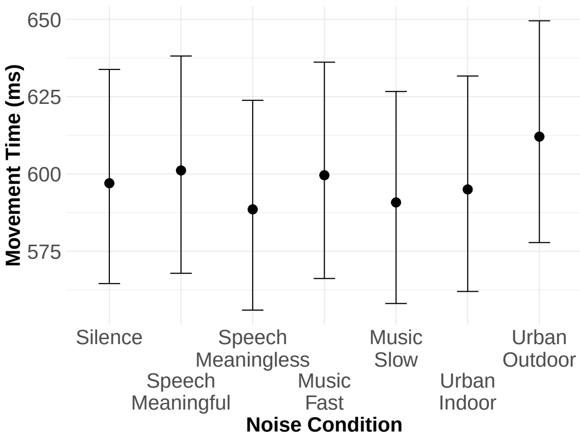
[69] R William Soukoreff and I Scott MacKenzie. 2004. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *International journal of human-computer studies* 61, 6 (2004), 751–789.
[70] Marco Speicher, Anna Maria Feit, Pascal Ziegler, and Antonio Krüger. 2018. Selection-based text entry in virtual reality. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.
[71] Stephen Stansfeld, John Gallacher, Wolfgang Babisch, and Martin Shipley. 1996. Road traffic noise and psychiatric disorder: prospective findings from the Caerphilly Study. *Bmj* 313, 7052 (1996), 266–267.
[72] John Sweller. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive science* 12, 2 (1988), 257–285.
[73] Garreth W Tigwell, Zhanna Sarsenbayeva, Benjamin M Gorman, David R Flatla, Jorge Goncalves, Yeliz Yesilada, and Jacob O Wobbrock. 2019. Addressing the challenges of situationally-induced impairments and disabilities in mobile interaction. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–8.
[74] Rens Van de Schoot, Sarah Depaoli, Ruth King, Bianca Kramer, Kaspar Märtens, Mahlet G Tadesse, Marina Vannucci, Andrew Gelman, Duco Veen, Joukje Willemsen, et al. 2021. Bayesian statistics and modelling. *Nature Reviews Methods Primers* 1, 1 (2021), 1.
[75] Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. 2021. Rank-normalization, folding, and localization: An improved Rˆ for assessing convergence of MCMC (with discussion). *Bayesian analysis* 16, 2 (2021), 667–718.
[76] N Venetjoki, A Kaarlela-Tuomaala, E Keskinen, and V Hongisto. 2006. The effect of speech and speech intelligibility on task performance. *Ergonomics* 49, 11 (2006), 1068–1091.
[77] Alan Traviss Welford. 1968. Fundamentals of skill. (1968).
[78] Jacob O Wobbrock. 2019. Situationally-induced impairments and disabilities. *Web Accessibility: A Foundation for Research* (2019), 59–92.
[79] Wen Woo Ee and Michihiko Kanachi. 2005. The Effects of Music Type and Volume on Short-Term Memory. *Tohoku Psychologica Folia* (2005).
[80] Robert S Woodworth and Edward Lee Thorndike. 1901. The influence of improvement in one mental function upon the efficiency of other functions.(I). *Psychological review* 8, 3 (1901), 247.
[81] Renjie Wu and Hsiang-Ting Chen. 2023. The effect of visual and auditory modality mismatching between distraction and warning on pedestrian street crossing behavior. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 1045–1054.
[82] Jing Yang, Amit Barde, and Mark Billinghurst. 2022. Audio augmented reality: A systematic review of technologies, applications, and future research directions. *journal of the audio engineering society* 70, 10 (2022), 788–809.
[83] Difeng Yu, Hai-Ning Liang, Xueshi Lu, Kaixuan Fan, and Barrett Ens. 2019. Modeling endpoint distribution of pointing selection tasks in virtual reality environments. *ACM Transactions on Graphics (TOG)* 38, 6 (2019), 1–13.
[84] Mingrui Ray Zhang, Shumin Zhai, and Jacob O Wobbrock. 2019. Text entry throughput: Towards unifying speed and accuracy in a single performance metric. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–13.
[85] Qiushi Zhou, Brandon Victor Syiem, Beier Li, Jorge Goncalves, and Eduardo Velloso. 2024. Reflected Reality: Augmented Reality through the Mirror. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 4 (2024), 1–28.
[86] Zhiying Zhou, Adrian David Cheok, Xubo Yang, and Yan Qiu. 2004. An experimental study on the role of 3D sound in augmented reality environment. *Interacting with Computers* 16, 6 (2004), 1043–1068.
[87] Joanne E Zucco, Bruce H Thomas, and Karen Grimmer. 2006. Evaluation of four wearable computer pointing devices for drag and drop tasks when stationary and walking. In *2006 10th IEEE International Symposium on Wearable Computers*. IEEE, 29–36.

A Integrated Loudness (LUFS) Values for the Audio Resources

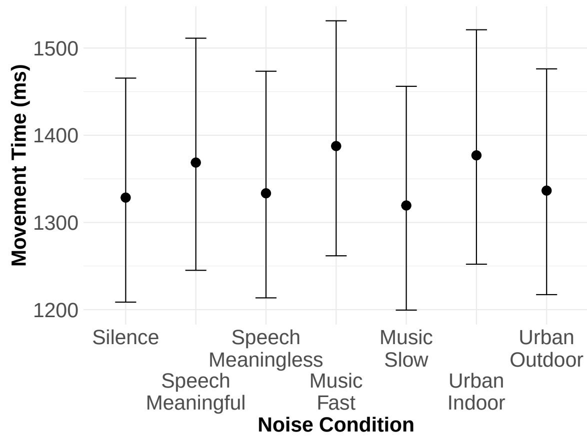
Table 4. Loudness measurements (LUFS) for all audio stimuli.

Stimulus	**Momentary Max**	**Short-Term Max**	**Integrated**
Meaningful Speech (Chinese)	-23.32	-26.55	-32.64
Meaningful Speech (English)	-21.50	-27.04	-30.85
Meaningful Speech (Japanese)	-22.20	-25.67	-30.61
Meaningful Speech (Kazakh)	-20.88	-26.03	-31.45
Fast Music	-8.03	-10.62	-14.95
Slow Music	-10.95	-13.81	-18.69
Urban Indoor	-23.32	-27.95	-35.30
Urban Outdoor	-13.41	-16.18	-20.68

B Visualisation of the Mean Values and 95% CI

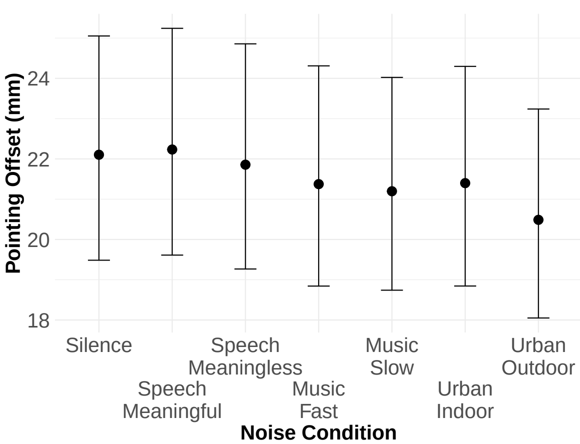


(a) Effect of noise on movement time under target acquisition (direct selection) task.

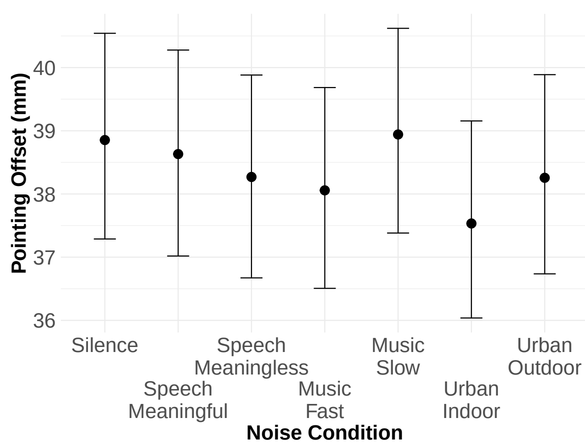


(b) Effect of noise on movement time under target acquisition (ray-casting) task.

Fig. 5. Model posterior predictions for movement time across different conditions of noise under two different tasks (direct selection and ray-casting). Scores correspond to the movement time in milliseconds (higher is worse). The upper bound and lower bound indicate the true value of the estimation lies within the 95% CI. The dot and center line are the predicted median movement time.

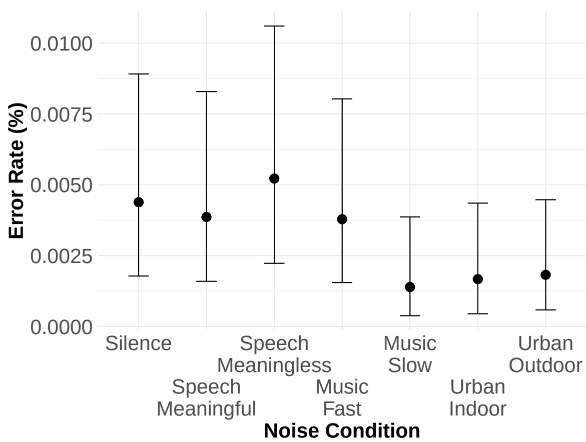


(a) Effect of noise on pointing offset under target acquisition (direct selection) task.

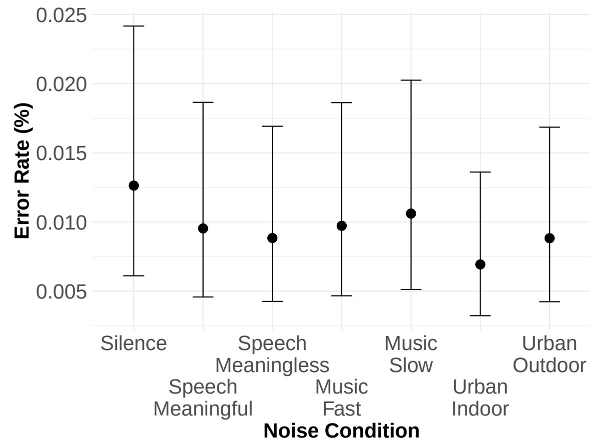


(b) Effect of noise on pointing offset under target acquisition (ray-casting) task.

Fig. 6. Model posterior predictions for pointing offset across different conditions of noise under two different tasks (direct selection and ray-casting). Scores correspond to the pointing offset in millimeters (higher is worse). The upper bound and lower bound indicate the true value of the estimation lies within the 95% CI. The dot and center line are the predicted median movement time.

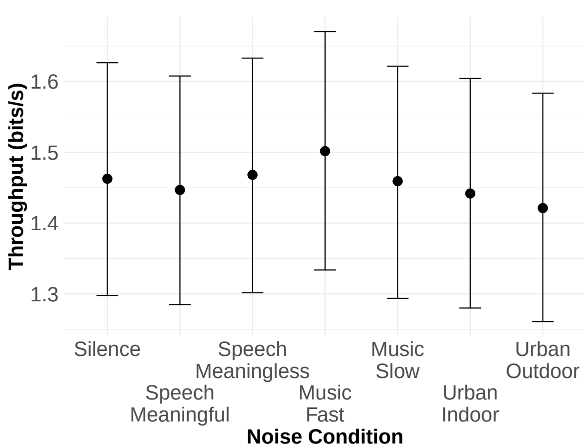


(a) Effect of noise on error rate under target acquisition (direct selection) task.

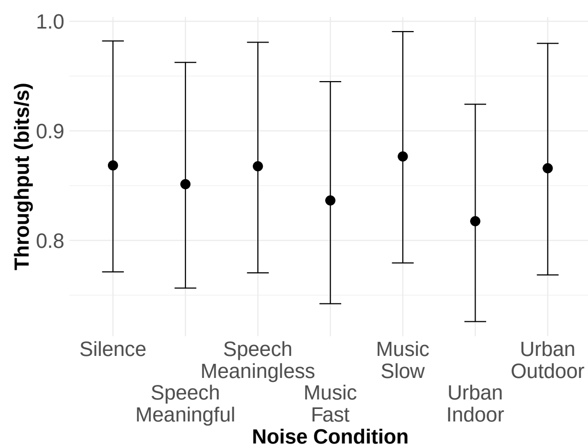


(b) Effect of noise on error rate under target acquisition (ray-casting) task.

Fig. 7. Model posterior predictions for error rate across different conditions of noise under two different tasks (direct selection and ray-casting). Scores correspond to the error rate in percentage (higher is worse). The upper bound and lower bound indicate the true value of the estimation lies within the 95% CI. The dot and center line are the predicted median movement time.

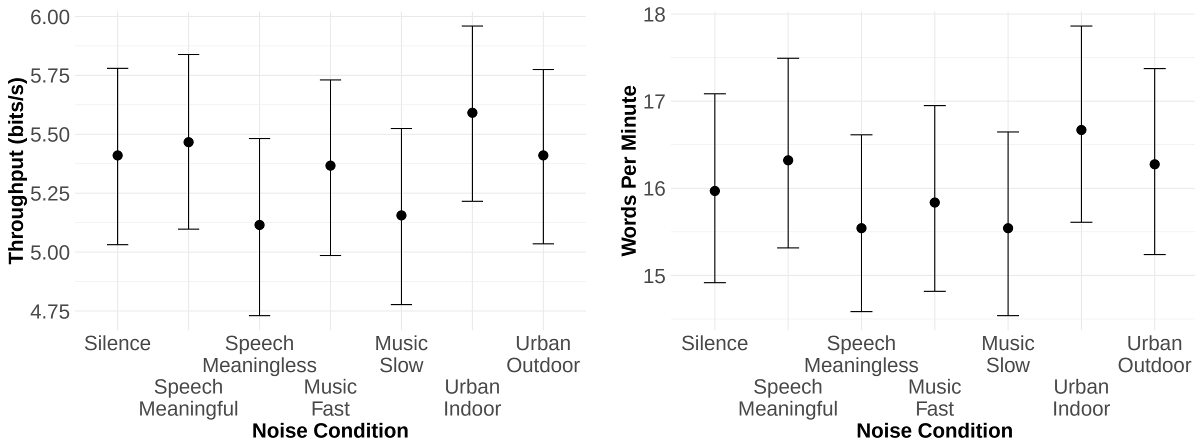


(a) Effect of noise on target acquisition throughput under target acquisition (direct selection) task.



(b) Effect of noise on target acquisition throughput under target acquisition (ray-casting) task.

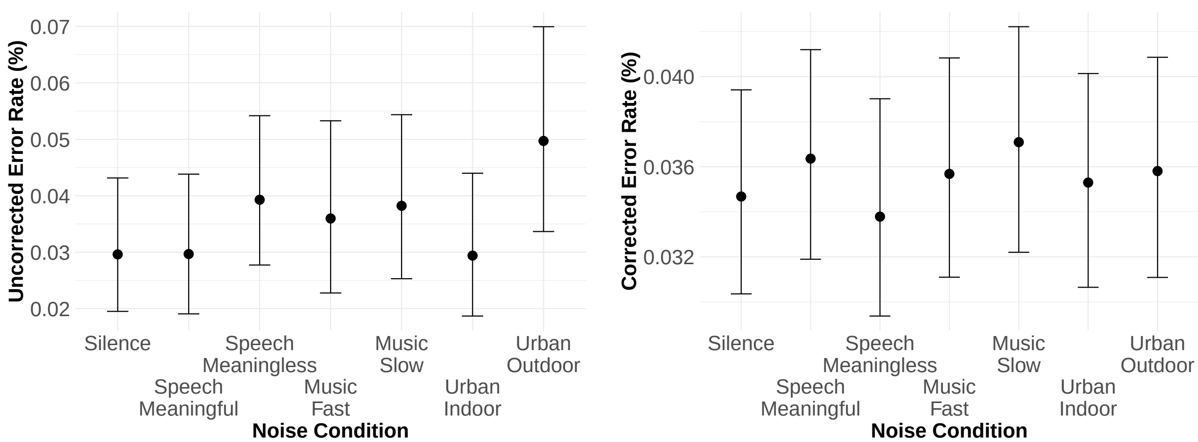
Fig. 8. Model posterior predictions for target acquisition throughput across different conditions of noise under two different tasks (direct selection and ray-casting). Scores correspond to the target acquisition throughput in bits/s (higher is worse). The upper bound and lower bound indicate the true value of the estimation lies within the 95% CI. The dot and center line are the predicted median movement time.



(a) Effect of noise on text entry throughput.

(b) Effect of noise on words per minute.

Fig. 9. Model posterior predictions for text entry throughput and words per minute across different conditions of noise. The upper bound and lower bound indicate the true value of the estimation lies within the 95% CI. The dot and center line are the predicted median movement time.



(a) Effect of noise on uncorrected error rate.

(b) Effect of noise on corrected error rate.

Fig. 10. Model posterior predictions for uncorrected and corrected error rate across different conditions of noise under text entry task. The upper bound and lower bound indicate the true value of the estimation lies within the 95% CI. The dot and center line are the predicted median movement time.

C Tables of the Complete Set of Parameter Estimates

Table 5. Summary of the movement time model: MT ∼ 1 + (1|participant) + order + condition · ID. We provide the posterior means of parameter estimates (Estimate), posterior error of these estimates (Error), the upper and lower bound of the 95% CI, and the one-sided posterior probability of a positive effect (PP). All parameter estimates converged with an ESS well above 1000 and an R-hat of 1.00.

Parameter	Direct Selection		Ray-Casting	
	Estimate (Error)	**95% CI**	**Estimate (Error)**	**95% CI**
Fixed Effects (Covariates)				
Intercept	5.95 (0.05)	[5.84, 6.05]	6.65 (0.07)	[6.52, 6.78]
Meaningful	−0.01 (0.05)	[−0.10, 0.08]	0.09 (0.05)	[−0.01, 0.19]
Meaningless	−0.05 (0.05)	[−0.14, 0.03]	0.04 (0.05)	[−0.06, 0.14]
Fast Music	0.06 (0.04)	[−0.03, 0.15]	0.01 (0.05)	[−0.09, 0.10]
Slow Music	−0.01 (0.04)	[−0.09, 0.08]	−0.03 (0.05)	[−0.13, 0.06]
Urban Indoor	0.09 (0.04)	[0.00, 0.18]	0.02 (0.05)	[−0.08, 0.12]
Urban Outdoor	0.08 (0.05)	[−0.01, 0.16]	0.01 (0.05)	[−0.09, 0.11]
Fixed Effects (Independent Variables)				
Order	−0.07 (0.00)	[−0.07, −0.07]	−0.03 (0.00)	[−0.04, −0.03]
ID	0.23 (0.04)	[0.16, 0.31]	0.56 (0.05)	[0.47, 0.65]
ID:Meaningful	0.02 (0.06)	[−0.09, 0.13]	−0.08 (0.06)	[−0.20, 0.05]
ID:Meaningless	0.04 (0.06)	[−0.07, 0.15]	−0.05 (0.06)	[−0.17, 0.08]
ID:Fast Music	−0.07 (0.06)	[−0.18, 0.04]	0.06 (0.06)	[−0.07, 0.18]
ID:Slow Music	−0.01 (0.06)	[−0.13, 0.09]	0.03 (0.06)	[−0.09, 0.16]
ID:Urban Indoor	−0.13 (0.06)	[−0.24, −0.02]	0.03 (0.06)	[−0.10, 0.16]
ID:Urban Outdoor	−0.05 (0.06)	[−0.16, 0.06]	−0.00 (0.06)	[−0.13, 0.12]
Random Effects				
Participant (SD)	0.22 (0.03)	[0.16, 0.30]	0.27 (0.04)	[0.20, 0.36]
Further Distributional Parameters				
sigma	0.37 (0.00)	[0.37, 0.38]	0.43 (0.00)	[0.42, 0.43]
ndt	197.98 (1.65)	[194.51, 200.96]	153.91 (2.44)	[148.81, 158.25]

Table 6. Summary of the pointing offset model: Offset ~ 1 + (1|participant) + order + condition · width. We provide the posterior means of parameter estimates (Estimate), their posterior errors (Error), and the lower and upper bounds of the 95% credible intervals (95% CI). All estimates converged with an ESS well above 1000 and R-hat values at or near 1.00.

Parameter	**Direct Selection**		**Ray-Casting**	
	Estimate (Error)	**95% CI**	**Estimate (Error)**	**95% CI**
Fixed Effects (Independent Variables)				
Intercept	2.44 (0.09)	[2.27, 2.62]	2.68 (0.06)	[2.57, 2.79]
Meaningful	−0.17 (0.09)	[−0.35, 0.00]	−0.04 (0.08)	[−0.19, 0.12]
Meaningless	−0.06 (0.09)	[−0.24, 0.12]	0.06 (0.08)	[−0.09, 0.22]
Fast Music	−0.14 (0.09)	[−0.32, 0.03]	0.09 (0.08)	[−0.06, 0.24]
Slow Music	−0.06 (0.09)	[−0.24, 0.11]	−0.02 (0.08)	[−0.18, 0.13]
Urban Indoor	−0.05 (0.09)	[−0.23, 0.12]	0.04 (0.08)	[−0.11, 0.19]
Urban Outdoor	−0.12 (0.09)	[−0.30, 0.06]	0.01 (0.08)	[−0.14, 0.17]
Fixed Effects (Covariates)				
Order	0.01 (0.00)	[0.00, 0.01]	0.00 (0.00)	[−0.00, 0.01]
Width	0.01 (0.00)	[0.00, 0.01]	0.01 (0.00)	[0.01, 0.01]
Width:Meaningful	0.00 (0.00)	[0.00, 0.01]	0.00 (0.00)	[−0.00, 0.00]
Width:Meaningless	0.00 (0.00)	[−0.00, 0.00]	−0.00 (0.00)	[−0.00, 0.00]
Width:Fast Music	0.00 (0.00)	[−0.00, 0.00]	−0.00 (0.00)	[−0.00, 0.00]
Width:Slow Music	0.00 (0.00)	[−0.00, 0.00]	0.00 (0.00)	[−0.00, 0.00]
Width:Urban Indoor	0.00 (0.00)	[−0.00, 0.00]	−0.00 (0.00)	[−0.00, 0.00]
Width:Urban Outdoor	0.00 (0.00)	[−0.00, 0.00]	−0.00 (0.00)	[−0.00, 0.00]
Random Effects				
Participant (SD)	0.31 (0.05)	[0.24, 0.42]	0.08 (0.01)	[0.05, 0.10]
Further Distributional Parameters				
sigma	0.69 (0.00)	[0.69, 0.70]	0.60 (0.00)	[0.59, 0.61]
ndt	0.01 (0.01)	[0.00, 0.03]	0.02 (0.02)	[0.00, 0.08]

Table 7. Summary of the error rate model: Error Rate $\sim$ 1 + (1|participant) + order + condition $\cdot$ ID. We provide the posterior means of parameter estimates (Estimate), their posterior errors (Error), and the lower and upper bounds of the 95% credible intervals (95% CI). All estimates converged with an ESS well above 1000 and R-hat values at or near 1.00.

Parameter	**Direct Selection**		**Ray-Casting**	
	Estimate (Error)	**95% CI**	**Estimate (Error)**	**95% CI**
Fixed Effects (Independent Variables)				
Intercept	−5.04 (1.05)	[−7.12, −3.02]	−4.26 (0.49)	[−5.25, −3.30]
Meaningful	0.04 (1.35)	[−2.55, 2.78]	−0.55 (0.50)	[−1.52, 0.44]
Meaningless	−1.24 (1.30)	[−3.80, 1.29]	−0.83 (0.54)	[−1.87, 0.24]
Fast Music	−1.64 (1.45)	[−4.45, 1.15]	−0.58 (0.52)	[−1.62, 0.45]
Slow Music	−3.29 (1.96)	[−7.34, 0.44]	0.19 (0.52)	[−0.85, 1.21]
Urban Indoor	0.82 (1.67)	[−2.45, 4.02]	−0.79 (0.60)	[−1.95, 0.43]
Urban Outdoor	−2.94 (1.76)	[−6.44, 0.42]	−0.74 (0.53)	[−1.77, 0.29]
Fixed Effects (Covariates)				
Order	0.26 (0.06)	[0.14, 0.39]	0.11 (0.03)	[0.06, 0.17]
ID	−1.57 (1.36)	[−4.32, 0.92]	−0.61 (0.47)	[−1.54, 0.32]
ID:Meaningful	−0.20 (1.88)	[−3.99, 3.43]	0.36 (0.64)	[−0.89, 1.60]
ID:Meaningless	1.85 (1.70)	[−1.42, 5.24]	0.61 (0.69)	[−0.74, 1.97]
ID:Fast Music	1.95 (1.87)	[−1.65, 5.62]	0.42 (0.67)	[−0.92, 1.76]
ID:Slow Music	2.79 (2.39)	[−1.97, 7.54]	−0.48 (0.70)	[−1.86, 0.91]
ID:Urban Indoor	−2.36 (2.52)	[−7.46, 2.41]	0.24 (0.78)	[−1.32, 1.77]
ID:Urban Outdoor	2.69 (2.16)	[−1.53, 6.98]	0.50 (0.68)	[−0.82, 1.83]
Random Effects				
Participant (SD)	1.17 (0.33)	[0.65, 1.93]	1.60 (0.28)	[1.15, 2.25]

Table 8. Summary of the target acquisition throughput: TP ~ 1 + (1|participant) + order + condition · width. We provide the posterior means of parameter estimates (Estimate), their posterior errors (Error), and the lower and upper bounds of the 95% credible intervals (95% CI). All estimates converged with an ESS well above 1000 and R-hat values at or near 1.00.

Parameter	Direct Selection		Ray-Casting	
	Estimate (Error)	**95% CI**	**Estimate (Error)**	**95% CI**
Fixed Effects (Independent Variables)				
Intercept	−0.65 (0.06)	[−0.77, −0.53]	−1.19 (0.07)	[−1.33, −1.06]
Meaningful	−0.09 (0.03)	[−0.14, −0.04]	0.01 (0.04)	[−0.08, 0.10]
Meaningless	−0.08 (0.03)	[−0.13, −0.03]	−0.01 (0.05)	[−0.10, 0.08]
Fast Music	0.01 (0.03)	[−0.05, 0.06]	−0.11 (0.05)	[−0.20, −0.02]
Slow Music	−0.05 (0.03)	[−0.11, −0.01]	−0.03 (0.05)	[−0.12, 0.06]
Urban Indoor	−0.09 (0.03)	[−0.14, −0.04]	−0.10 (0.04)	[−0.19, −0.01]
Urban Outdoor	−0.15 (0.03)	[−0.20, −0.10]	−0.01 (0.05)	[−0.10, 0.08]
Fixed Effects (Covariates)				
Order	0.04 (0.00)	[0.04, 0.04]	0.04 (0.00)	[0.03, 0.04]
IDe	0.95 (0.02)	[0.91, 0.99]	0.84 (0.03)	[0.78, 0.90]
IDe:Meaningful	0.08 (0.03)	[0.03, 0.14]	−0.03 (0.04)	[−0.11, 0.05]
IDe:Meaningless	0.09 (0.03)	[0.04, 0.14]	0.01 (0.04)	[−0.08, 0.09]
IDe:Fast Music	0.02 (0.03)	[−0.04, 0.07]	0.07 (0.04)	[−0.02, 0.16]
IDe:Slow Music	0.06 (0.03)	[0.01, 0.11]	0.04 (0.04)	[−0.05, 0.12]
IDe:Urban Indoor	0.08 (0.03)	[0.03, 0.13]	0.04 (0.04)	[−0.04, 0.12]
IDe:Urban Outdoor	0.13 (0.03)	[0.08, 0.19]	0.01 (0.05)	[−0.08, 0.10]
Random Effects				
Participant (SD)	0.29 (0.04)	[0.22, 0.39]	0.31 (0.05)	[0.23, 0.41]
Further Distributional Parameters				
sigma	0.28 (0.00)	[0.28, 0.29]	0.48 (0.00)	[0.47, 0.48]
ndt	0.00 (0.00)	[0.00, 0.00]	0.00 (0.00)	[0.00, 0.00]

Table 9. Summary of the text entry throughput: TP ∼ 1 + (1|participant) + order + condition. We provide the posterior means of parameter estimates (Estimate), posterior error of these estimates (Error), the upper and lower bound of the 95% CI, and the one-sided posterior probability of a positive effect (PP). All parameter estimates converged with an ESS well above 1000 and an R-hat of 1.00.

Parameter	**Estimate (Error)**	**95% CI**
Fixed Effects (Independent Variables)		
Intercept	4.75 (0.20)	[4.36, 5.15]
Meaningful	0.06 (0.14)	[−0.22, 0.33]
Meaningless	−0.30 (0.14)	[−0.57, −0.02]
Fast Music	−0.04 (0.15)	[−0.33, 0.24]
Slow Music	−0.25 (0.14)	[−0.54, 0.03]
Urban Indoor	0.18 (0.15)	[−0.11, 0.47]
Urban Outdoor	0.00 (0.15)	[−0.28, 0.29]
Fixed Effects (Covariates)		
Order	0.22 (0.02)	[0.18, 0.26]
Random Effects		
Participant (SD)	0.76 (0.12)	[0.57, 1.04]
Further Distributional Parameters		
sigma	1.17 (0.03)	[1.11, 1.22]

Table 10. Summary of the uncorrected error rate and corrected error rate: UER/CER ~ 1 + (1|participant) + order + condition · width. We provide the posterior means of parameter estimates (Estimate), posterior error of these estimates (Error), the upper and lower bound of the 95% CI, and the one-sided posterior probability of a positive effect (PP). All parameter estimates converged with an ESS well above 1000 and an R-hat of 1.00.

Parameter	**UER**		**CER**	
	Estimate (Error)	**95% CI**	**Estimate (Error)**	**95% CI**
Fixed Effects (Independent Variables)				
Intercept	-1.69 (0.27)	[-2.22, -1.19]	-2.74 (0.08)	[-2.90, -2.59]
Meaningful	0.00 (0.31)	[-0.62, 0.62]	0.05 (0.09)	[-0.12, 0.22]
Meaningless	0.35 (0.29)	[-0.21, 0.92]	-0.03 (0.09)	[-0.21, 0.16]
Fast Music	0.24 (0.33)	[-0.40, 0.88]	0.03 (0.09)	[-0.15, 0.21]
Slow Music	0.31 (0.30)	[-0.28, 0.91]	0.07 (0.09)	[-0.10, 0.25]
Urban Indoor	-0.00 (0.32)	[-0.63, 0.60]	0.02 (0.09)	[-0.16, 0.19]
Urban Outdoor	0.66 (0.32)	[0.03, 1.28]	0.03 (0.09)	[-0.15, 0.22]
Fixed Effects (Covariates)				
Order	0.01 (0.04)	[-0.08, 0.09]	0.01 (0.01)	[-0.01, 0.03]
Random Effects				
Participant (SD)	0.11 (0.08)	[0.00, 0.31]	0.05 (0.03)	[0.00, 0.12]
Further Distributional Parameters				
phi	2.68 (0.30)	[2.11,3.31]	49.20 (3.21)	[43.03, 55.55]
zi	0.81 (0.01)	[0.79, 0.84]	0.44 (0.02)	[0.41, 0.47]

Table 11. Summary of the words per minute: WPM ~ 1 + (1|participant) + order + condition. We provide the posterior means of parameter estimates (Estimate), posterior error of these estimates (Error), the upper and lower bound of the 95% CI, and the one-sided posterior probability of a positive effect (PP). All parameter estimates converged with an ESS well above 1000 and an R-hat of 1.00.

Parameter	**Estimate (Error)**	**95% CI**
Fixed Effects (Independent Variables)		
Intercept	2.62 (0.04)	[2.55, 2.69]
Meaningful	0.02 (0.03)	[−0.03, 0.08]
Meaningless	−0.03 (0.03)	[−0.08, 0.03]
Fast Music	−0.01 (0.03)	[−0.06, 0.05]
Slow Music	−0.03 (0.03)	[−0.08, 0.03]
Urban Indoor	0.04 (0.03)	[−0.01, 0.10]
Urban Outdoor	0.02 (0.03)	[−0.03, 0.07]
Fixed Effects (Covariates)		
Order	0.04 (0.00)	[0.03, 0.05]
Random Effects		
Participant (SD)	0.14 (0.02)	[0.10, 0.19]
Further Distributional Parameters		
sigma	0.22 (0.01)	[0.21, 0.23]
ndt	0.03 (0.03)	[0.00, 0.12]