

How Much Delay Goes Unnoticed? Quantifying Perceptual Latency Thresholds Across Multimodalities in Virtual Reality

Natthaphon Hatsakornkhanachok^[0009–0006–7206–2763]
and Parinya Punpongsanon^[0000–0003–2720–7768]

Saitama University, Saitama 338-8570, Japan
`hatsakornkhanachok.n.637@ms.saitama-u.ac.jp`

Abstract. In virtual reality (VR), delays between a user’s action and sensory feedback disrupt immersion and responsiveness, yet careful budgeting can expand rather than degrade VR’s design. We quantify perceptual latency thresholds across three modalities (lighting, sound, rotation) using a fixed-staircase procedure with ternary judgments ($N = 20$) to estimate the Point of Subjective Simultaneity (PSS) and Temporal Binding Window (TBW). Sound showed the smallest unnoticed delay, rotation the largest, and thresholds differed significantly by modality. Extending to a conventional display revealed no platform differences, indicating thresholds are governed by sensorimotor engagement rather than immersion, yielding latency guidelines for VR application.

Keywords: Multisensory Integration · Perceptual Latency · Temporal Perception.

1 Introduction

Virtual Reality (VR) enables the creation of simulated environments that incorporate visual, auditory, and other sensory feedback. Specifically, VR generates immersive three-dimensional environments to foster interactive experiences between humans and systems, such as in gaming or training scenarios [17, 20, 21]. When a user presses a virtual button, any temporal discrepancy between the input and the sensory consequence can disrupt the sense of presence and diminish perceived system responsiveness. However, in gamification or training simulation, achieving effectively zero latency across multi-modalities remains practically challenging. Furthermore, these feedback channels naturally have different delays because the software engine processes and renders each signal differently. Prior research has extensively explored multisensory integration, particularly between audio-visual and visual-tactile modalities such as rapid recalibration driven by visual-intensity-dependent processing time [7] and examined the effect of perceptual training on the audio-visual [14]. Although VR provides greater immersion than conventional displays, both environments rely on a fundamental sensorimotor engagement loop: users perform an action and expect immediate feedback. Therefore, an important question is whether perceptual latency

thresholds are determined by the common sensorimotor process underlying active interaction. However, whether these perceptual latency thresholds remain consistent in HMDs and extend to conventional displays during active interaction has not been systematically investigated. To address this question, we quantify perceptual latency thresholds across three sensory modalities. Participants actively pressed a button to trigger a sensory event while latency was manipulated using a fixed-staircase procedure with ternary judgments to estimate the PSS and the TBW. Although auditory-visual temporal asymmetry is established [18], [25], threshold values for interactive VR design remain uncharacterized. Similarly, whether immersion affects this sensitivity [8], [12] has not been directly tested against a conventional display using an identical task; we address this gap through quantitative thresholds and a direct platform comparison. This study isolates the effect of display platform (HMD vs. conventional display) under an identical virtual interaction paradigm, in which both button activation and sensory feedback are digitally rendered without physical or mechanical actuation. By evaluating identical interaction tasks in HMD and extending the study to a conventional display, our work quantifies perceptual latency thresholds that provide latency guidelines for designing responsive gamified experiences and training simulations.

In summary, the main contributions of this paper are as follows:

- We virtual realsystematically quantify perceptual latency thresholds across three sensory modalities in HMD.
- We extend the study to conventional displays by applying the same virtual environment and experimental procedures.
- Through our experiments, we show that perceptual latency thresholds differ significantly by modality.
- We derive latency guidelines that help developers prioritize sensory feedback and develop responsive applications, such as gamified experiences and training simulations.

2 Related Work

Our work builds upon multisensory integration and simultaneity thresholds, temporal perception and time perception in virtual reality, and HMDs and conventional displays.

2.1 Multisensory Integration and Simultaneity Thresholds

In everyday interactions, users expect near-immediate feedback following an action, such as pressing a button to activate a light. When the delay between an action and its outcome exceeds a specific threshold, the interaction is no longer perceived as instantaneous, but rather as temporally disjointed [2]. Consistent with this, latency degrades both user preference and task performance in direct-touch interaction, with effects observed even below 10 ms and worsening

for smaller or more distant targets [9, 15]. These thresholds are closely tied to temporal sensitivity, which varies across sensory modalities [25], so acceptable latency depends significantly on the feedback modality involved.

A key framework for understanding these limits relies on the concepts of PSS and TBW. The PSS represents the temporal offset at which two events are most likely to be perceived as simultaneous, while the TBW defines the temporal interval within which asynchronous events are perceptually integrated as a single event [24]. Crucially, PSS and TBW capture two different dimensions of temporal perception. The PSS identifies the perceptual center, revealing which sensory modality the brain naturally expects to experience first [18], whereas TBW maps the maximum binding window for causality [25] and is known to narrow with age [1]. Notably, active interaction such as pressing a virtual button alters temporal expectations through the sense of agency. Building on this, we examine lighting, sound, and rotation to quantify perceptual latency thresholds, systematically determining latency guidelines for gamified experience and training simulation.

2.2 Temporal Perception vs. Time Perception

Temporal perception and time perception are closely related but represent distinct cognitive processes. Temporal perception concerns the detection and integration of events occurring over short timescales, typically ranging from milliseconds to a few hundred milliseconds, whereas time perception refers to the subjective estimation of duration over longer intervals [16, 22]. Human time processing is adaptive, and a more time-precise sense, such as hearing, can shift the perceived timing of a less precise sense, such as vision [13, 18], and humans continuously integrate multisensory information within specific temporal boundaries to preserve causality and synchrony [27]. Prior work on time perception has largely focused on explicit duration judgments and higher-level factors such as attention and cognitive load [4, 5], as well as factors such as visual motion [11], task complexity [23], and user engagement [6], rather than the low-level temporal characteristics of sensory feedback itself.

By quantifying the boundaries within which delays go unnoticed, gamification or training simulation may manipulate temporal dynamics without disrupting users' perception of responsiveness or causality.

2.3 HMDs vs. Conventional Displays

HMDs and conventional displays provide fundamentally different interaction environments. HMDs provide an immersive experience through stereoscopic display, FOV, and interactive elements. Conversely, conventional displays typically offer more limited sensory engagement, and these differences influence user experience, attention, and task performance [5]. Prior improvement on head-tracking latency thresholds has focused primarily on HMDs [8, 12], and richer FOV and depth cues in such devices could plausibly increase sensitivity to temporal inconsistencies relative to a conventional display [19]. However, no prior study

has directly quantified perceptual latency thresholds across both platforms under an identical task, leaving it unexplored whether PSS and TBW generalize across display technologies. To address this gap, we quantified perceptual latency thresholds in HMD and extended the study to a conventional display using an identical interaction paradigm and sensory conditions.

3 Experiments

We conducted two experiments. The first experiment quantified perceptual latency thresholds across three sensory modalities, including lighting, sound, and rotation in HMD. The second experiment examined whether these perceptual latency thresholds remain consistent when the same tasks are performed on a conventional display.

3.1 Experimental Setup

The experimental setup consists of two separate environments, i.e., VR and a conventional display. While they share the same virtual scene, their implementations are separate.

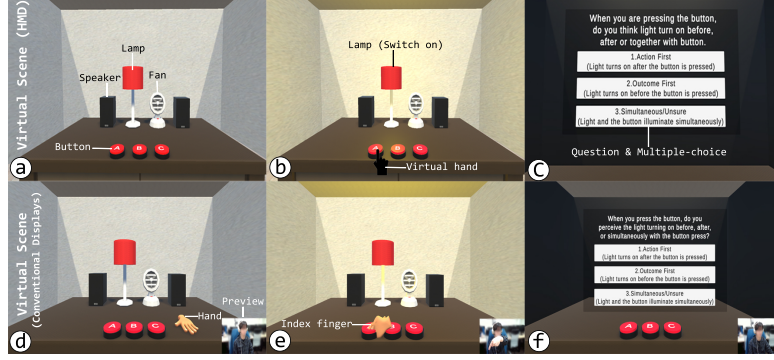


Fig. 1. VR and conventional display experimental setup; (a) The virtual scene (HMD) featuring a speaker, a lamp, a fan, and buttons. (b) In the virtual scene, when each participant presses a button using their index finger, the lamp activates and remains illuminated for approximately 1 second. (c) The system then displays the question and multiple-choice answers within the virtual scene. Each participant provides their responses verbally to the experimenter, who records the data via a keyboard. (d) The virtual scene (conventional display) featuring a speaker, a lamp, a fan, a hand, buttons, and a user preview. (e)-(f) The experimental procedure is similar to that in (b) and (c).

Implementation and User Interaction The virtual environment was developed using Unity (2022.3.47f1) with the Oculus Integration package and the Low Poly Office Set (1.1.0) asset pack. Within this simulated scene (Fig 1a), users interact with four main virtual elements designed for intuitive interaction: three buttons, a lamp (lighting), a speaker (sound), and a fan (rotation). For VR, participants used a Meta Quest 3 with integrated hand tracking (Fig 1b). For the conventional display, an Intel RealSense D415 depth camera with MediaPipe (0.10.3) and Simple Hands Prefab (1.0) tracked the virtual hand (Fig 1d). In both platforms, the right index finger pressed the virtual button to trigger feedback.

Stimuli Generation and Calibration For the VR setup, the lighting modality utilized a lamp-enclosed point light (range: 12, intensity: 1, yellow hue) rendered via real-time lighting for immediate feedback. During the conventional display, the sound modality played an MP3 file at 60% volume through noise-canceling headphones (Sony WH-1000XM2). The rotation modality simulated a fan rotating at 800 RPM, a speed identified during preliminary testing as optimal for clear visibility. For the conventional display setup, the sound and rotation parameters matched the VR condition. However, we adjusted the lighting configuration (range: 50, intensity: 18.5, yellow hue) while maintaining real-time lighting rendering to ensure cross-platform visual and temporal consistency on the monitor. We acknowledge that the simulated fan includes a brief acceleration phase before reaching full rotational speed, mirroring the behavior of a physical fan. While this design choice was intended to enhance realism, it may introduce an additional perceptual delay independent of the manipulated latency parameter, potentially inflating the measured PSS for the motion condition. We discuss this further as a limitation in Section 7. Latency was measured over 50 test rounds per condition. In VR (72 Hz), lighting and rotation latency was < 0.01 ms and sound latency was 14 ms; the latter was subtracted. In the conventional display (75 Hz), measured latencies were 21 ms for lighting and rotation and 31 ms for sound, which were subtracted for temporal calibration.

Participants and Procedure We recruited 20 participants (aged between 20 and 30, 16 male and 4 female, all right-handed) to participate in our experiment. They agreed to voluntarily participate without receiving compensation. Before starting the experiment, participants scanned the QR code displayed on the screen to access the online consent form. Once they had completed the form, we explained the experimental procedure. Note that all participants took part in both experiment 1 and experiment 2 in the same order. Participants sat on a chair and used the right index finger to press a virtual button. The selected feedback was then presented for approximately 1 second; after completing the modality, they viewed a multiple-choice question within the scene and responded to the experimenter.

Task We quantified delay thresholds ranging from -400 ms to 400 ms across all three modalities. Employing a fixed staircase procedure with a ternary judgment task, the system latency was adjusted by $+10$ ms when participants perceived Outcome first or Simultaneous synchrony, and by -10 ms for an Action-first response. We note that our staircase procedure groups ‘Outcome first’ and ‘Simultaneous’ responses to drive the $+10$ ms step, effectively reducing the ternary judgment to a two-alternative criterion that converges toward the stimulus level at which the probability of an ‘Action-first’ response falls to approximately 50%. Modality order was randomized and counterbalanced across participants. Each participant completed 150 trials total (3 modalities $\times$ 50 trials), with a 1-minute rest period provided between modalities. The entire experiment required approximately 45 minutes to complete. The task is identical in both experiment 1 and experiment 2.

3.2 Statistical Analysis

To analyze the experimental results, we employed statistical procedures, including data truncation, the PSS formula based on the arithmetic mean, and parametric estimation formulas for TBW.

Data Truncation Each experimental condition consisted of 50 trials. Following standard fixed staircase protocols [10], the first 15 trials were excluded as an initial burn-in phase characterized by large fluctuations near distant starting points [3].

The Point of Subjective Simultaneity (PSS) We calculated the arithmetic mean of the stimulus levels from the stable tracking phase (trials 16 to 50) to determine the PSS value as follows.

$$PSS = \frac{1}{n} \sum_{i=16}^{50} L_i \quad (1)$$

In Equation 1, L_i represents the delay level (ms) at trial i and n denotes the total number of analyzed trials ($n = 35$). Additionally, we computed a reversal-based PSS, the mean SOA level at staircase reversal points, occurring within the stable tracking phase (trials 16 – 50). We report the starting SOA, mean reversal counts, and this reversal-based estimate alongside the mean-based PSS in Section 4.1 and Section 5.1.

We calculated the standard deviation σ as follows:

$$\sigma = \sqrt{\frac{\sum_{i=16}^{50} (L_i - PSS)^2}{n - 1}} \quad (2)$$

In Equation 2, L_i represents the delay in each trial (spanning 16 to 50), PSS represents the mean across these trials for each participant, n the total number

of analyzed trials ($n = 35$, where $n - 1 = 34$) and $\sum$ represents the sum of squared differences.

The Temporal Binding Window (TBW) We calculated the precision of temporal integration as follows.

$$TBW = (z_{0.75} - z_{0.25}) \times \sigma \quad (3)$$

In Equation 3, σ represents the standard deviation of the delay levels from trial 16 to 50, and z_p denotes the standard score (z-score) at the p^{th} percentile of a normal distribution. Specifically, $(z_{0.75} - z_{0.25})$ equals approximately 1.349. It should be noted that σ here reflects the dispersion of the staircase levels during the stable tracking phase rather than the slope of a fitted psychometric function.

4 Experiment 1: HMD

In this experiment, we quantify the perceptual latency thresholds across three sensory modalities within a virtual environment. The study examines how temporal delays affect users' ability to detect multisensory asynchrony. In addition, we formulated the following hypothesis for our study.

- H1-1: PSS for the sound modality will be lower than that of the lighting and rotation modalities.
- H1-2: TBW for the rotation modality will be broader than that for lighting and sound modalities.

This hypothesis aligns with basic principles of how the brain processes various types of sensory feedback [28]. Regarding the PSS, the brain processes auditory signals significantly faster than visual (lighting) and kinesthetic (rotation) inputs [18]. This expectation arises because users typically perceive auditory feedback almost instantly after initiating an action, such as a button press. Regarding the TBW, the temporal window for grouping distinct senses depends heavily on the signal type. While lighting and sound function as transient, instantaneous events, rotation (e.g., a rotating fan) is a continuous event spanning a prolonged duration [25]. Because tracking rotation requires the central nervous system to process ongoing movement rather than a discrete onset timestamp, the perceptual system naturally accommodates a wider temporal window to bind this rotation to the user's action [26].

4.1 Result

As shown in Figure 2, the PSS results across the three modalities were 154.83 ± 18.34 ms (lighting), 95.12 ± 21.60 ms (sound), and 164.70 ± 16.54 ms (rotation), respectively (all values are reported as mean $\pm$ SEM). The staircase for all participants began at a fixed starting SOA of 0 ms, with an average reversal-stable count of 13.60 for lighting, 14.80 for sound, and 14.20 for rotation. The

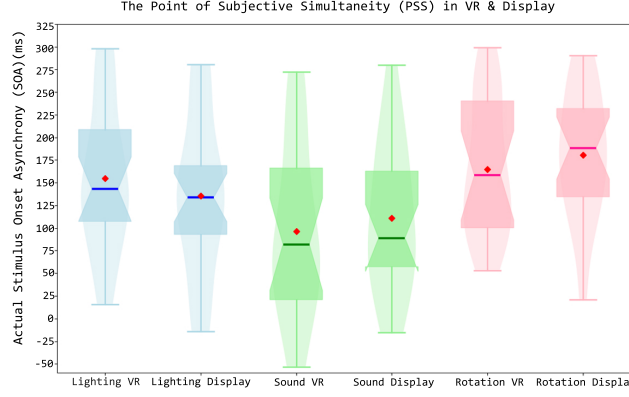


Fig. 2. The results of PSS in VR and conventional display included 3 modalities: lighting, sound, and rotation.

corresponding reversal-based PSS estimates were 164.23 ± 23.02 ms for lighting, 94.73 ± 24.85 ms for sound, and 163.30 ± 18.37 ms for rotation. A one-way RM-ANOVA was conducted to assess whether there were significant differences in PSS across these modalities. The results indicated a significant difference ($F(2,38) = 17.73$, $p < 0.01$). The post hoc analysis with a Bonferroni correction revealed that the PSS for the sound modality was significantly different from both the lighting ($p = 0.01$) and rotation ($p < 0.01$) modalities.

As shown in Figure 3, the TBW results were 50.84 ± 30.25 ms (lighting), 43.05 ± 21.71 ms (sound) and 48.33 ± 21.36 ms (rotation). A one-way RM-ANOVA revealed no statistically significant difference between the modalities ($F(2,38) = 1.48$, $p = 0.23$).

4.2 Discussion

Our results demonstrate that the PSS differed significantly across the three modalities; no significant differences were observed in TBW within the virtual environment. These findings validate hypothesis H1-1 and reject H1-2.

Regarding H1-1, the PSS distribution varied distinctly across the three modalities. In the lighting modality, immediate visual feedback from the illuminated table allowed rapid, confident responses. Conversely, the sound modality induced perceptual confusion, as participants noted a subtle discrepancy immediately following the button press. Similar judgment challenges occurred in the rotation modality; participants experienced temporal uncertainty regarding the exact onset of the fan's rotation.

Regarding H1-2, the TBW analysis indicated that the rotation modality did not exhibit a significantly greater width compared to lighting and sound, thus rejecting the initial hypothesis. For context, delays in real-world interconnected devices such as Wi-Fi- or Zigbee-controlled lighting are generally understood to

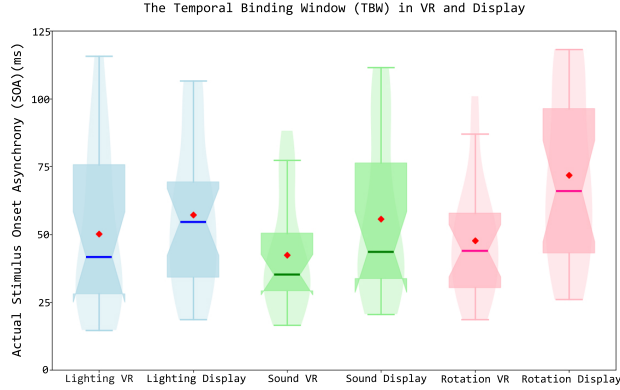


Fig. 3. The results of TBW in VR and conventional display included 3 modalities: lighting, sound, and rotation.

be substantially longer than the imperceptibility thresholds reported here, often due to communication overhead such as device wake-up and wireless transmission latency. This suggests that our thresholds represent a considerably stricter standard than typical smart-home latency: rather than identifying the point at which a delay becomes merely tolerable, our thresholds mark the point at which it becomes entirely unnoticed. Additionally, physical devices provide direct tactile feedback (such as the physical click of a button), which is absent in our virtual setup and may change how users expect and tolerate delays. Directly comparing physical and virtual interactions, both with and without tactile feedback, would be an interesting direction for future research.

Interestingly, rotation showed higher PSS values than lighting, indicating a descriptive tendency for latency to be detected at a higher delay for rotation. However, the difference in TBW between the two modalities was not statistically significant. Therefore, the present results do not provide sufficient evidence to conclude that participants were more precise in judging latency for rotation than for lighting.

5 Experiment 2: Conventional Display

In this experiment, we examined whether these perceptual latency thresholds remain consistent when the same tasks are performed on a conventional display. In addition, we formulated the following hypothesis for our study.

- H2-1: The Point of Subjective Simultaneity (PSS) for the sound modality will be lower than that of the lighting and rotation modalities.
- H2-2: The Temporal Binding Windows (TBW) for the rotation modality will be broader than those for lighting and sound modalities.
- H2-3: The PSS in VR will be lower, and the TBW will be narrower than that of the conventional display.

5.1 Result

We conducted a two-way RM-ANOVA to examine the effects of platform (VR vs. conventional display) and modality (lighting, sound, and rotation). As shown in Fig 2, the PSS analysis revealed no significant main effect of platform ($F(1,19) = 0.03$, $p = 0.86$). However, the main effect of modality was significant ($F(2,38) = 34.20$, $p < 0.01$), and there was no significant interaction between platform and modality ($F(2,38) = 2.92$, $p = 0.06$).

Regarding the TBW, Fig 3 shows no significant main effect of platform ($F(1,19) = 0.56$, $p = 0.46$). However, the main effect of modality was significant ($F(2,38) = 22.39$, $p < 0.01$), and there was no significant interaction between platform and modality ($F(2,38) = 2.30$, $p = 0.11$).

As shown in Figure 2, the PSS results across the three modalities were 134.36 ± 19.89 ms (lighting), 109.57 ± 18.49 ms (sound) and 179.73 ± 15.69 ms (rotation) for conventional display. As in Experiment 1, the staircase began with a fixed starting SOA of 0 ms, with an average reversal-stable count of 13.60 for lighting, 14.60 for sound, and 11.65 for rotation. The corresponding reversal-based PSS estimates were 137.36 ± 22.16 ms for lighting, 115.60 ± 21.82 ms for sound, and 196.52 ± 23.57 ms for rotation. A one-way RM-ANOVA was conducted to assess whether there were significant differences in PSS across these modalities. The results indicated a significant difference ($F(2,38) = 17.17$, $p < 0.01$). The post hoc analysis with a Bonferroni correction revealed that the PSS for the sound modality was significantly different from both the lighting ($p = 0.03$) and rotation ($p = 0.01$) modalities.

As shown in Figure 3, the TBW results were 57.30 ± 28.11 ms (lighting), 55.62 ± 29.25 ms (sound) and 72.20 ± 31.75 ms (rotation) for conventional display. A one-way RM-ANOVA revealed a statistically significant difference among the modalities ($F(2,38) = 6.56$, $p < 0.04$). The post-hoc analysis with a Bonferroni correction revealed that the TBW in the sound modality was significantly different from that in the rotation modality ($p = 0.05$), whereas no significant difference was found compared to the lighting modality ($p = 1.00$).

5.2 Discussion

Our results demonstrate that the PSS differed significantly across the three modalities; a significant difference in TBW was observed between the sound and rotation modalities, but no significant difference between the lighting and rotation modalities in a conventional display. These findings confirm the latency noticeability thresholds, thereby validating hypotheses H2-1 and H2-2, while rejecting H2-3.

Regarding H2-1, the PSS distributions across the three modalities revealed distinct perceptual patterns. In the lighting modality, immediate visual feedback from the illuminated table facilitated rapid, confident responses. Conversely, the sound modality induced perceptual confusion due to a subtle onset discrepancy. The rotation modality presented similar challenges in temporal judgment. Because participants primarily focused on the button-pressing task, simultaneously

tracking the fan’s rotation caused visual attention splits and response uncertainty.

Regarding H2-2, the TBW analysis confirmed that the rotation modality exhibited a greater width than both lighting and sound modalities. We note that the discrepancy between experiments regarding the rotation TBW H1-2 was rejected in the HMD condition, whereas H2-2 was supported on the conventional display.” First, we completed the HMD before the conventional display conditions (see Section 7); increased familiarity or fatigue by Experiment 2 may have altered participants’ sensitivity to the continuous rotation modality stimulus independently of the conventional display itself. Second, an HMD has richer depth cues and a wider field of view to support more precise visual tracking of the fan’s rotation than the conventional display, potentially narrowing the effective TBW for motion specifically in the HMD.

Finally, regarding H2-3, our two-way RM-ANOVA showed no significant main effect of platform and no significant interaction between platform and modality. This null result is consistent with the interpretation that latency thresholds are governed primarily by sensorimotor engagement rather than platform-specific immersion. We note that participants completed a practice session before data collection and that we randomized and counterbalanced modality order, mitigating (though not fully eliminating) concerns about learning effects. However, because platform order itself was fixed, we present the platform-invariance conclusion as consistent with, rather than conclusively confirming, this interpretation, and recommend counterbalanced platform ordering in future replications.

6 Latency Guidelines

We define the latency guideline as the PSS value itself, representing the empirically estimated point at which action and feedback are perceived as simultaneous. However, we report TBW separately as a comparative dispersion measure indicating each modality and platform’s relative sensitivity to timing variability, rather than combining it with PSS to form a tolerance interval as follows.

$$\textit{Latency Guidelines} = \textit{PSS} \tag{4}$$

We summarize latency guidelines for maintaining imperceptible delays during interaction for general users, as shown in Table 1.

7 Limitation

Our investigations have five limitations.

First, the HMD and conventional display platforms differed in several confounded respects beyond immersion itself: a minor refresh-rate discrepancy (72 vs. 75 Hz, unlikely to matter relative to our 10 ms step size), FOV and stereoscopic depth cues present only in the HMD, greater susceptibility to environmental distraction on the conventional display, and a non-identical physical interaction posture (natural button press in VR vs. an extended-arm depth-camera

Modality	Platform	Latency Guideline (PSS) (ms)	TBW (ms) Comparative Only
Lighting	VR	155 ms	50.84 ms
Sound	VR	95 ms	43.05 ms
Rotation	VR	165 ms	48.33 ms
Lighting	Display	134 ms	57.30 ms
Sound	Display	110 ms	55.62 ms
Rotation	Display	180 ms	72.20 ms

Table 1. Recommended Latency Guidelines Derived from PSS, but TBW should not be interpreted as a validated absolute tolerance width.

posture on the conventional display). Future work will use unified hand tracking to equalize posture across platforms

Second, our rotation stimulus (a virtual fan with a brief acceleration phase) mirrors realistic fan behavior, but this realism may itself add a perceptual delay independent of the manipulated latency, potentially confounding the measured thresholds. Future work should isolate motion-specific perception using instantaneous-onset stimuli or kinesthetic/force feedback devices.

Third, platform order was fixed, while modality order was randomized. A practice session and rest breaks mitigated but did not fully rule out learning/fatigue effects on the observed platform invariance (H2-3) and the rotation TBW discrepancy between Experiments 1 and 2 (Section 5.2). Future work should counterbalance platform order to isolate this effect.

Fourth, our study is limited to virtual-to-virtual interaction and does not compare against physical, mechanically actuated devices; such comparisons, with and without tactile confirmation, remain important future work.

Finally, despite 50 trials per condition maintaining internal reliability, the limited sample size may have inflated inter-subject variability; future studies should scale up participant numbers to strengthen statistical power and clarify the PSS-TBW correlation.

8 Conclusion

In this paper, we quantified perceptual latency thresholds across lighting, sound, and rotation modalities in interactive environments. By employing a ternary-judgment staircase task, we estimated the PSS and TBW, revealing that sound required the smallest unnoticed delay and rotation the largest, with thresholds differing significantly by modality. Comparing virtual reality with a conventional display showed no significant platform differences, indicating that these temporal thresholds are governed primarily by sensorimotor engagement rather than display immersion. Ultimately, these findings establish foundational latency guidelines for optimizing responsive applications.

Acknowledgments. This work has been partially supported by JSPS Fellows Grant Number 16J00049 and JST BOOST Grant Number JPMJBY25E2.

References

1. Ainsworth, K., Bertone, A.: Audiovisual temporal binding window narrows with age in autistic individuals. *Autism Research* **16**(2), 355–363 (2023). <https://doi.org/10.1002/aur.2860>
2. Eagleman, D.M.: Human time perception and its illusions. *Current Opinion in Neurobiology* **18**(2), 131–136 (2008). <https://doi.org/10.1016/j.conb.2008.06.002>, cognitive neuroscience
3. Gescheider, G.A.: *Psychophysics: the fundamentals*. Psychology Press (1997), <https://doi.org/10.4324/9780203774458>, discusses the stabilization of threshold estimates in adaptive procedures.
4. Gibbon, J.: Scalar expectancy theory and weber’s law in animal timing. *Psychological Review* **84**(3), 279–325 (1977), <https://psycnet.apa.org/doi/10.1037/0033-295X.84.3.279>
5. Grondin, S.: Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception, & Psychophysics* **72**(3), 561–582 (2010). <https://doi.org/10.3758/APP.72.3.561>
6. Harjunen, V.J., Spapé, M., Ravaja, N.: Anticipation of aversive visual stimuli lengthens perceived temporal duration. *Psychological Research* **86**(4), 1230–1238 (jun 2022). <https://doi.org/10.1007/s00426-021-01559-6>
7. Horsfall, R., Harrison, N., Meyer, G., Wuerger, S.: Perceived audio-visual simultaneity is recalibrated by the visual intensity of the preceding trial. *Multisensory Research* **37**(2), 143 – 162 (2024). <https://doi.org/10.1163/22134808-bja10121>
8. Jerald, J., Whitton, M.: Relating scene-motion thresholds to latency thresholds for head-mounted displays. In: 2009 IEEE Virtual Reality Conference. pp. 211–218 (2009). <https://doi.org/10.1109/VR.2009.4811025>
9. Jota, R., Ng, A., Dietz, P., Wigdor, D.: How fast is fast enough? a study of the effects of latency in direct-touch pointing tasks. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. p. 2291–2300. CHI ’13, Association for Computing Machinery, New York, NY, USA (2013). <https://doi.org/10.1145/2470654.2481317>
10. Levitt, H.: Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America* **49**(2B), 467–477 (1971), <https://doi.org/10.1121/1.1912375>
11. Li, J., Kim, J.E.: The effect of task complexity on time estimation in the virtual reality environment: An eeg study. *Applied Sciences* **11**(20) (2021). <https://doi.org/10.3390/app11209779>
12. Mania, K., Adelstein, B.D., Ellis, S.R., Hill, M.I.: Perceptual sensitivity to head tracking latency in virtual environments with varying degrees of scene complexity. In: Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization. p. 39–47. APGV ’04, Association for Computing Machinery, New York, NY, USA (2004). <https://doi.org/10.1145/1012551.1012559>
13. Marucci, M., Di Flumeri, G., Borghini, G., Sciaraffa, N., Scandola, M., Pavone, E.F., Babiloni, F., Betti, V., Aricò, P.: The impact of multisensory integration and perceptual load in virtual reality settings on performance, workload and presence. *Scientific Reports* **11**(1), 4831 (mar 2021). <https://doi.org/10.1038/s41598-021-84196-8>
14. McGovern, D.P., Burns, S., Hirst, R.J., Newell, F.N.: Perceptual training narrows the temporal binding window of audiovisual integration in both younger and older adults. *Neuropsychologia* **173**, 108309 (2022). <https://doi.org/10.1016/j.neuropsychologia.2022.108309>

15. Ng, A., Lepinski, J., Wigdor, D., Sanders, S., Dietz, P.: Designing for low-latency direct-touch input. In: *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*. p. 453–464. UIST '12, Association for Computing Machinery, New York, NY, USA (2012). <https://doi.org/10.1145/2380116.2380174>
16. Pöppel, E.: A hierarchical model of temporal perception. *Trends in Cognitive Sciences* **1**(2), 56–61 (1997). [https://doi.org/10.1016/S1364-6613\(97\)01008-5](https://doi.org/10.1016/S1364-6613(97)01008-5)
17. Song, J., Yang, K., Yang, J.: The application of virtual reality in games. In: *2022 IEEE 2nd International Conference on Data Science and Computer Application (ICDSCA)*. pp. 1086–1090 (Oct 2022). <https://doi.org/10.1109/ICDSCA56264.2022.9988576>
18. Spence, C., Squire, S.: Multisensory integration: Maintaining the perception of synchrony. *Current Biology* **13**(13), R519–R521 (2003). [https://doi.org/10.1016/S0960-9822\(03\)00445-7](https://doi.org/10.1016/S0960-9822(03)00445-7)
19. Strasburger, H., Rentschler, I., Jüttner, M.: Peripheral vision and pattern recognition: A review. *Journal of Vision* **11**(5), 13–13 (2011), <https://doi.org/10.1167/11.5.13>
20. Thanyadit, S., Punpongsanon, P., Piumsomboon, T., Pong, T.C.: Xr-live: Enhancing asynchronous shared-space demonstrations with spatial-temporal assistive toolsets for effective learning in immersive virtual laboratories. *Proc. ACM Hum.-Comput. Interact.* **6**(CSCW1) (Apr 2022), <https://doi.org/10.1145/3512983>
21. Thanyadit, S., Punpongsanon, P., Pong, T.C.: Efficient information sharing techniques between workers of heterogeneous tasks in 3d cve. *Proc. ACM Hum.-Comput. Interact.* **2**(CSCW) (Nov 2018), <https://doi.org/10.1145/3274441>
22. Tinker, M.A.: Temporal perception. In: Boring, E.G., Langfeld, H.S., Weld, H.P. (eds.) *Psychology: A factual textbook*, pp. 246–259. John Wiley & Sons, Inc., New York (1935), <https://psycnet.apa.org/doi/10.1037/13517-010>
23. Tonoyan, Y., Fornaciai, M., Parsons, B., Bueti, D.: Subjective time is predicted by local and early visual processing. *NeuroImage* **264**, 119707 (2022). <https://doi.org/10.1016/j.neuroimage.2022.119707>
24. van Wassenhove, V., Grant, K.W., Poeppel, D.: Temporal window of integration in auditory-visual speech perception. *Neuropsychologia* **45**(3), 598–607 (2007), <https://doi.org/10.1016/j.neuropsychologia.2006.01.001>, advances in Multisensory Processes
25. Vroomen, J., Keetels, M.: Perception of intersensory synchrony: A tutorial review. *Attention, Perception, & Psychophysics* **72**(4), 871–884 (may 2010). <https://doi.org/10.3758/APP.72.4.871>
26. Wallace, M.T., Stevenson, R.A.: The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia* **64**, 105–123 (2014). <https://doi.org/10.1016/j.neuropsychologia.2014.08.005>
27. Wallace, M.T., Stevenson, R.A.: Multisensory temporal integration: task and stimulus dependencies. *Experimental Brain Research* **232**(4), 1115–1126 (2014). <https://doi.org/10.1007/s00221-013-3507-3>
28. Welch, R.B., Warren, D.H.: Immediate perceptual response to intersensory discordance. *Psychological Bulletin* **88**(3), 638–663 (1980). <https://doi.org/10.1037/0033-2909.88.3.638>