

Perceptual Sensitivity to Network Latency for Guidance Scenarios in Co-Located AR Collaboration

Froso Sarri^{1,*†}, Katerina Mania^{1,†}

¹*School of Electrical and Computer Engineering, Technical University of Crete, Greece*

Abstract

Understanding how system latency affects perception is critical, particularly in networked collaborative Extended Reality (XR) systems. Rather than just measuring system latency, task performance, or deploying usability questionnaires, it is essential that latency perception be measured based on formal psychophysical methods in specific settings. The work proposed aims to investigate sensitivity thresholds to system latency in networked collaborative XR environments, utilized in a simplified upper limb rehabilitation scenario. It seeks to determine the perceptual thresholds of latency that users can tolerate before task efficiency or coordination is compromised. We investigate latency perception in the context of a collaborative AR guidance paradigm designed to support referential communication. Findings establish acceptable latency bounds for networked collaborative applications, in a healthcare context.

Keywords

Human-Computer Interaction (HCI), Mixed/Augmented Reality, Psychophysics, XR Collaboration

1. Introduction

Augmented Reality (AR)-enabled collaboration can enhance contextual guidance in shared environments by providing clear, unambiguous cues that support faster and more accurate task completion [1]. In this context, understanding how system latency affects perception is critical, particularly in networked collaborative XR systems, as the effectiveness of such systems strongly depends on the temporal alignment of shared multimodal signals [2], [3]. Rather than just measuring system latency, task performance, or deploying usability questionnaires, it is essential that latency perception be measured based on formal psychophysical methods. Investigating perceptual thresholds of latency in collaborative XR environments in a specific context allows designers to determine the delays that users can tolerate without impairing performance or coordination, informing the development of XR systems that maintain seamless interaction and shared situational awareness across participants.

Latency refers to the time lag between a user's action and the corresponding system response, arising from input sensing, computation, rendering, and data transmission. Elevated latency negatively affects manual performance, prompting users to slow their actions and rely on "move-and-wait" strategies [4], [5]. While limited sensorimotor adaptation can occur when task preview is provided, temporal delay hampers adaptation to other forms of display distortion [6]. Prior research further indicates that latency in XR systems diminishes both user performance and perceived simulation realism [7], with particularly pronounced effects in 3D object placement tasks [8]. Because latency introduced by sensing, computation, rendering, and transmission is inherent to XR and teleoperation systems, it is essential to understand its perceptual detectability and underlying mechanisms in order to develop effective mitigation approaches, such as predictive compensation [9].

Psychophysics is a subfield of perceptual psychology that employs controlled behavioral methods to study the relationship between the physical magnitude of a stimulus and the sensation perceived by a human or animal observer [10]. The minimum change in a stimulus required to elicit a perceptible

Co-Living between Reality and Virtuality as a Daily Routine, CHI 2026 Workshop, April 14, 2026, Barcelona, Spain

*Corresponding author.

†These authors contributed equally.

✉ esarri@tuc.gr (F. Sarri); amania@tuc.gr (K. Mania)

ORCID 0009-0004-7341-739X (F. Sarri); 0000-0002-9457-5117 (K. Mania)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

difference in sensation is known as the difference threshold. In contemporary usage, the terms difference threshold and Just Noticeable Difference (JND) are frequently used interchangeably. Thresholds are commonly estimated by presenting stimuli to observers and asking whether they can detect them. Since biological systems such as humans exhibit non-deterministic responses, these thresholds are evaluated using statistical methods.

Latency is even more critical in XR for medical purposes. In the context of European project SUN -Social and Human-Centred XR [11], innovative technologies and models that enhance social interaction and immersive perception have been developed, by innovations in wearable sensors [12], haptic interfaces [13] wireless and high-performance streaming solutions. SUN explored innovative ways to connect physical objects and digital counterparts, creating complete rehabilitation scenarios tested in hospital settings for efficient rehabilitation therapy [14]. Developed solutions focused on enhancing upper limb rehabilitation through an AR environment that enables patients to perform goal-oriented exercises, such as pick-and-place tasks supported by haptic feedback for posture correction. The system includes a collaborative feature where the therapist can join the same AR environment using a second headset, allowing them to view the same virtual objects as the patient and provide real-time physical and visual guidance (Fig. 1).

This study aims to investigate sensitivity thresholds to system latency in networked collaborative XR environments, in a simplified experimental scenario analogous to SUN's rehabilitation scenario. Specifically, it seeks to determine the perceptual thresholds of latency that users can tolerate before task efficiency or coordination is compromised, and to explore how different types of AR cues modulate sensitivity to these delays. We investigate latency perception in the context of a collaborative AR guidance paradigm designed to support referential communication and coordinated action as tested in an upper limb rehabilitation scenario. The paradigm, as described in [15], follows a reference-action sequence framework [16], in which two users interact within a shared virtual environment: an instructor selects objects and specifies target locations using gaze-based cues and annotations, while a user performs pick-and-place actions on the indicated objects. The system incorporates multimodal feedback, including visualized gaze behaviour, annotation cues, and object manipulation events, to provide non-verbal guidance and facilitate disambiguation during task execution. Because the paradigm relies on the real-time exchange of gaze and action signals, its effectiveness can be compromised by temporal delays that may disrupt the correspondence between instruction and response.

To ensure accurate collaborative interactions under networked conditions, it is essential to understand users' sensitivity to temporal asynchrony within the system. However, the full collaborative setup [15] involves complex interactions between gaze, gestures, and object manipulation, which introduce variability due to user input which can interfere with the isolated effects of latency. To address this challenge, we conducted a controlled user study to measure perceptual thresholds of latency using a stripped-down version of the collaborative paradigm. By reducing the interaction to a minimal stimulus-response sequence, we systematically manipulated artificial latency while preserving the shared environment and network configuration, allowing us to directly assess users' ability to detect temporal discrepancies. This controlled evaluation provides perceptual limits on latency perception, which seeks to quantify the level of delay an XR system can exhibit without being perceptible to the user, and can be used to interpret performance and user experience in the complete collaborative system.

2. Methodology

The collaborative task engages two users within a shared virtual environment, each equipped with the same Head-Mounted-Display (Microsoft's Hololens 2) (Fig. 1). One user is responsible for manipulating 3D objects, picking them from a varied set arranged on a horizontal surface and placing them onto designated virtual shelves. Their actions are guided by the second user, who serves as the instructor, directing which objects to select and specifying their target locations.

With the aim to identify the smallest asynchrony that a user can detect, we conducted experiments using a controlled setup based on our configuration, in order to study the relation between stimulus



Figure 1: AR task space from the instructor's POV.

intensity and sensation reported by a human. The basic procedure for measuring thresholds is to present a stimulus to observers, artificial latency in this case, and asking them to report whether they perceive events as synchronous or not [7].

With the aim of isolating temporal perception, we modified our collaborative pick-and-place paradigm into a minimal procedure to eliminate variability introduced by instructor gestures or gaze movements. Additionally, the user must know when the stimulus sequence is initiated. Thus, we kept the same network configuration and shared environment, but replaced the instructor's actions with a button-press trigger. The experimental procedure goes as follows: both users are connected to the same network, sharing the same environment. The instructor initiates the process, while the other user acts as the observer. The former presses a button that triggers an audio stimulus on the observer's device. An artificial latency is then added between the audio stimulus and the cup movement. The observer's task is to indicate whether the latency is detectable. Due to the time-sensitive nature of the experiment, an audio cue was preferred over a visual cue, as auditory perception has higher temporal resolution [17]. To ensure that only the artificial latency is measured, independent of network latency, the event sequence occurs entirely on the observer's side. Specifically, the button press event is sent over the network to trigger the audio stimulus on the observer's device, and all subsequent events are handled locally.

The users both wear the same HMD. The two devices are connected on the same local network via WiFi, with a Peer-to-Peer (P2P) parameterized network topology with one device acting as a host and the other as a client. Networking logic is handled by Unity's NGO (Netcode for GameObjects) [18]. The alignment of the two devices' coordinate systems is implemented via QR code tracking, establishing a common coordinate system at the code's real-world location.

We used the 2-down/1-up adaptive staircase method for changing artificial latency's (stimulus intensity) value [19]. The algorithm decreases the latency by one step after two consecutive delay detection responses, and increases the latency by one step after each single "no" response [20]. After internal testing, our range was between 0 (always undetected) and 800ms (always detected). We began each staircase at 400ms, the midpoint of the range. Our initial step was 100ms, then we decreased it to 50ms after two reversals (change in the direction of the stimulus adjustment, from increasing to decreasing latency or vice-versa) and to 20ms after six reversals, for rapid identification of threshold regions and fine sampling towards the end. The procedure terminates after 12 reversals or 80 trials, whichever occurs first. For each participant's experiment, we obtained the detection rate of each unique delay value. We then fitted a four-parameter cumulative Gaussian to compute detection probabilities:

$$\Psi(l) = \gamma + (1 - \gamma - \lambda) \Phi\left(\frac{l - \mu}{\sigma}\right) \quad (1)$$

where l is the latency value, γ is the guess rate, λ the lapse rate, μ the PSE (50% detection point), σ the slope parameter (standard deviation), and Φ the standard normal CDF. The extracted thresholds are the PSE (Point of Subjective Equality) at which an observer is equally likely to report latency or not (50% detection probability), and the JND, which is the additional latency required to raise detection probability from 50% to 75%. For a cumulative-Gaussian psychometric function with a standard-deviation parameter σ , JND is defined as:

$$\text{JND} = \sigma \Phi^{-1}(0.75) \approx 0.6745 \sigma. \quad (2)$$

3. Results

The experiment included 5 participants (1F, 4M). We consider this an acceptable number of users in psychophysics as a human sensory system varies much less between individuals than complex cognitive traits. Figure 2, displays the psychometric functions (cumulative Gaussian) for each participant. The extracted psychophysical threshold parameters PSE and JND can be seen in Table 1. PSE mean is 322.84 ms ($\sigma=103$ ms) and JND mean is 66.24 ms ($\sigma=27.86$ ms). The PSE value at 322.84 ms indicates the latency at which participants were equally likely to report the presence or absence of latency, reflecting a subjective point of equality rather than a detection threshold. The JND (66.24 ms) reflects the sensitivity of participants to changes in latency, with smaller values indicating steeper psychometric functions and higher discrimination sensitivity.

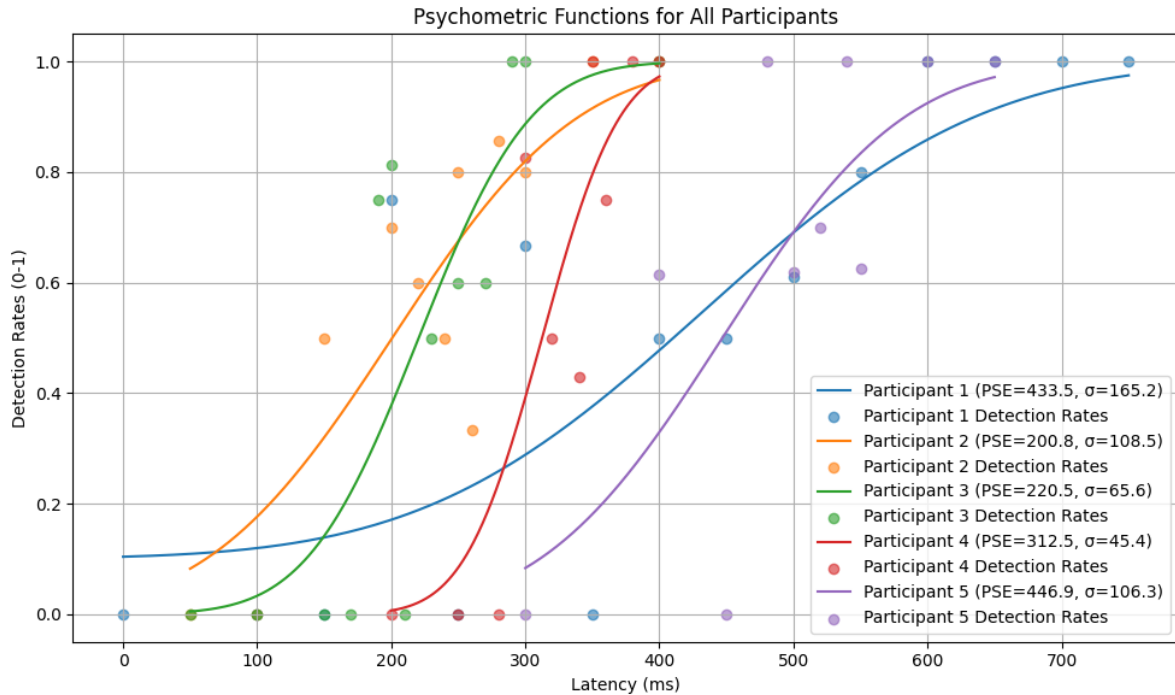


Figure 2: The psychometric functions for each participant.

4. Discussion

The ultimate goal in this research domain is to inform predictive latency compensation strategies enhancing performance and user experience. Latencies substantially above the PSE of 322.84 ms (e.g.,

Table 1

Extracted PSE and JND values for each participant.

Participant	PSE (ms)	JND (ms)
1	433.5	111.5
2	200.8	73.2
3	220.5	44.2
4	312.5	30.6
5	446.9	71.7

PSE + JND of 66.24 ms) were detected with higher probability, whereas latencies below this range were more likely to go unnoticed. Psychometric detection is probabilistic and there is no sharp threshold as detection probability increases gradually, not suddenly. These findings establish temporal limits that can inform system design requirements and acceptable latency bounds for networked collaborative applications. As such, these should be assessed closely to specific contexts and domains, not necessarily universally, as psychophysics establishes clear, unambiguous cause-and-effect relationships for sensory mechanisms before these being studied, in more complex, ecologically valid settings. In particular, knowledge of perceptual thresholds provides a quantitative target for latency masking techniques, allowing the collaborative experience to be seamless even under fluctuating network conditions. For future work, it is worth considering how individual differences, such as sensorimotor skills or prior XR experience, influence latency perception and task outcomes in collaborative AR/XR scenarios. This is crucial, especially for XR systems for medical purposes.

Acknowledgments

This work has received financial support by the Horizon Europe Research & Innovation Programme under Grant agreement N. 101092612 (Social and hUman ceNtered XR - SUN project).

Declaration on Generative AI

During the preparation of this work, the author(s) used Gemini in order to: Grammar and spelling check. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

- [1] S. Feng, W. He, X. Zhang, M. Billingham, S. Wang, A comprehensive survey on ar-enabled local collaboration, *Virtual Reality* 27 (2023) 2941–2966.
- [2] Y. Liang, G. Li, G. Zhang, J. Guo, Q. Liu, J. Zheng, T. Wang, Latency reduction in immersive systems through request scheduling with digital twin networks in collaborative edge computing, *ACM Transactions on Sensor Networks* (2024).
- [3] S. Van Damme, J. Sameri, S. Schwarzmann, Q. Wei, R. Trivisonno, F. De Turck, M. Torres Vega, Impact of latency on qoe, performance, and collaboration in interactive multi-user virtual reality, *Applied Sciences* 14 (2024) 2290.
- [4] I. S. MacKenzie, C. Ware, Lag as a determinant of human performance in interactive systems, in: *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, 1993, pp. 488–493.
- [5] C. Jay, M. Glencross, R. Hubbold, Modeling the effects of delayed haptic and visual feedback in a collaborative virtual environment, *ACM Transactions on Computer-Human Interaction (TOCHI)* 14 (2007) 8–es.

- [6] A. Khasawneh, H. Rogers, J. Bertrand, K. C. Madathil, A. Gramopadhye, Human adaptation to latency in teleoperated multi-robot human-agent search and rescue teams, *Automation in Construction* 99 (2019) 265–277.
- [7] K. Mania, B. D. Adelstein, S. R. Ellis, M. I. Hill, 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*, 2004, pp. 39–47.
- [8] B. Watson, N. Walker, P. Woytiuk, W. Ribarsky, Maintaining usability during 3d placement despite delay, in: *IEEE Virtual Reality, 2003. Proceedings.*, IEEE, 2003, pp. 133–140.
- [9] T. J. Buker, D. A. Vincenzi, J. E. Deaton, The effect of apparent latency on simulator sickness while using a see-through helmet-mounted display: Reducing apparent latency with predictive compensation, *Human factors* 54 (2012) 235–249.
- [10] G. A. Gescheider, *Psychophysics: the fundamentals*, Psychology Press, 2013.
- [11] L. Greci, F. Bosco, V. Croce, The social and human centered xr: Sun xr project, in: *International Conference on Extended Reality*, Springer, 2023, pp. 223–231.
- [12] P. Kasnesis, T. Plavoukou, A. C. Syropoulou, L. Toumanidis, G. Georgoudis, A knee rehabilitation exercises dataset for postural assessment using wearable devices, *Scientific Data* 12 (2025) 610.
- [13] M. Efraimidis, K. Mania, Wireless haptic glove using nitinol as a force feedback actuator, in: *Proceedings of the 2025 ACM International Symposium on Wearable Computers*, 2025, pp. 216–219.
- [14] F. Sarri, P. Kasnesis, S. Symeonidis, I. T. Paraskevopoulos, S. Diplaris, F. Posteraro, G. Georgoudis, K. Mania, Embodied augmented reality for lower limb rehabilitation (2024).
- [15] F. Sarri, K. Mania, Effect of gaze visualization on task efficiency and user behavior for guidance scenarios in co-located ar collaboration, in: *2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, IEEE, 2025, pp. 561–568.
- [16] S. Andrist, M. Gleicher, B. Mutlu, Looking coordinated: Bidirectional gaze mechanisms for collaborative interaction with virtual characters, in: *Proceedings of the 2017 CHI conference on human factors in computing systems*, 2017, pp. 2571–2582.
- [17] P. S. Zélanti, S. Droit-Volet, Auditory and visual differences in time perception? an investigation from a developmental perspective with neuropsychological tests, *Journal of experimental child psychology* 112 (2012) 296–311.
- [18] Netcode for gameobjects, 2026. Available at: <https://docs-multiplayer.unity3d.com/netcode/current/about/>. Accessed: 2026-02-09.
- [19] H. Levitt, Transformed up-down methods in psychoacoustics, *The Journal of the Acoustical society of America* 49 (1971) 467–477.
- [20] B. Treutwein, Adaptive psychophysical procedures, *Vision research* 35 (1995) 2503–2522.