ABC2: Journal of Architecture, Building, Construction, and Cities

Empowering research for Sustainable Development Goals, ABC2: Architecture, Building, Construction, and Cities is a fundamental manifesto to address these pressing issues, fostering dialogue and knowledge exchange among researchers, practitioners, and policymakers. Exploring sustainable design, resilient infrastructure, advanced construction methods, and equitable urban development, ABC2 aims to empower the global community to create adaptive, inclusive, and sustainable environments. The ABC2 focus on cutting-edge research, technological advancements, and transformative strategies is essential for navigating the future of our cities and communities.

Research Article

Gesture-based Survey Design for Evaluating Indoor Spatial Experience in Extended Reality

Hyeokjin Choi[1], Jae Yeon Shin[1], Choo Sheriel Shannon[1], Seung Hyun Cha[1]*

[1*] Korea Advanced Institute of Science and Technology, Daejeon, South Korea

DOI: https://doi.org/10.66408/abc2.2025.15

Correspondence: shcha@kaist.ac.kr

Received: 15/05/2025
Revised: 04/09/2025
Accepted: 09/12/2025
Published: 13/12/2025

Volume: 2025
Issue: 01
Pages: 60-67

Abstract

The integration of Extended Reality (XR) into built environments presents opportunities for spatial experience, yet evaluation methods fail to capture unique characteristics when physical and virtual elements coexist. Current approaches—from usability-focused to architecture-oriented methods—overlook verbal-behavioral discrepancies where users report positive experiences while exhibiting avoidance behaviors. To address this gap, we present SEEMEE (Spatial Experience Evaluation Method for Extended Environments), a framework evaluating XR spatial experience through behavioral congruence measurement—the degree to which user movements match natural behaviors in equivalent physical spaces. SEEMEE employs a three-stage guideline capturing behavioral responses to spatial stimuli. Target Movement Setting establishes baselines by documenting Personal Pattern Baseline (individual movement characteristics) and Spatial Pattern Baseline (navigation trajectories within environmental constraints). Affection Panoply Modeling captures experiential responses at multiple temporal scales: fleeting affection within 3-second windows and savoring affection at 5-minute, 30-minute, and 24-hour intervals. Response Congruence Formatting interprets behavioral differences through Temporal Distillation Protocol (tracking adaptation patterns) and Uncertainty Distillation Protocol (distinguishing system-level issues from individual preferences). To detail, we demonstrate a virtual window experiment as an example. Findings reveal that SEEMEE effectively captures temporal patterns of verbal-behavioral congruence. Positive self-reports paired with avoidance trajectories, or naturalistic movement aligned with consistent self-reports, provide composite indicators that conventional questionnaires miss. The proposed guidelines offer new possibilities for evaluating genuine spatial experience.

Keywords: Extended reality; Spatial experience; Qualitative assessment; Human-Space interaction

Highlights

- SEEMEE evaluates XR spatial experience through behavioral congruence measurement.
- SEEMEE enable systematic assessment of physical-virtual behavioral differences.
- SEEMEE captures verbal-behavioral discrepancies that traditional questionnaires miss.

1 Introduction

Extended Reality (XR) technologies are increasingly deployed in indoor environments, from architectural visualisation to retail experiences and workplace collaboration. As these applications mature beyond technical demonstrations, evaluating how users experience XR-augmented spaces becomes critical. However, current evaluation methods often fail to capture the unique nuance of spatial experience when physical and virtual elements coexist. This paper presents SEEMEE (Spatial Experience Evaluation Method for Extended Environments), a survey design framework that evaluates XR spatial experience through behavioural congruence—the degree to which user movements in XR environments match their behaviours in equivalent physical spaces.

Spatial experience in XR differs fundamentally from traditional digital interactions. Users navigate physical rooms while encountering virtual objects, creating complex embodied experiences that standard usability metrics cannot adequately assess. Existing evaluation tools fall into three cases. Conventional usability focused instruments like the System Usability Scale (Brooke, 1996) and NASA-TLX (Hart & Staveland, 1988) measure interface quality and cognitive load but overlook spatial behaviours. Architecture-oriented methods like Post Occupancy Evaluation (Preiser, White, & Rabinowitz, 2015) assess physical spaces but cannot account for virtual elements. Application-specific methods designed to verify psychological effects have been proposed and can be useful in certain contexts, but they often prove inadequate for measuring spatial experiences with limited functional requirements (Halbig & Latoschik, 2021). Commonly, these approaches rely heavily on self-reports, while measurements through physiological signals often risk leading to overly reductive interpretations. Most importantly, these methods fail to capture genuine spatial experience as revealed through unconscious behavioural responses.

The core challenge is that users often report positive experiences while exhibiting avoidance behaviours. For instance, users may rate a virtual window as "natural" yet maintain twice the normal viewing distance. This verbal-behavioural discrepancy might suggest that conscious evaluation and embodied response diverge in XR spaces. Traditional questionnaires cannot capture this divergence because they focus on subjective satisfaction rather than behavioural naturalness. Furthermore, the temporal dynamics of XR experience—from immediate reactions to consolidated memories—require multi-layered assessment approaches that current tools lack.

To address this research gap, we propose that movement patterns serve as primary indicators of XR spatial experience quality. User movement patterns serve as critical indicators revealing the degree of verbal–behavioural congruence. For instance, avoidance trajectories co-occurring with positive self-reports, or naturalistic spatial behaviours aligned with consistent verbal evaluations, both provide essential cues for interpreting the authenticity of XR experience. This behavioural congruence approach enables multi-layered interpretation of the relationship between reported and enacted responses, thereby offering objective criteria for assessing the quality of spatial experience in XR environments. Our design guideline addresses these challenges by capturing nuanced behavioural responses to spatial stimuli for interpreting authentic spatial experience.

2 Survey Design Guideline

The SEEMEE framework consists of three sequential stages for designing surveys that evaluate spatial experience in XR environments. The following is a visualisation overview to help understand the proposed concepts (see Fig. 1).

Each stage produces specific outputs that inform subsequent stages. **Target Movement Setting** establishes baselines for comparing physical and virtual spatial behaviours. This stage identifies which movements to observe and defines their normal ranges. Two complementary approaches are employed: (i) On-Body Movement Selection produces the Personal Pattern Baseline, documenting individual movement characteristics such as head rotation and gesture patterns in physical space; (ii)

Off-Body Movement Selection generates the Spatial Pattern Baseline, capturing navigation trajectories and positioning preferences within environmental constraints. **Affection Panoply Modelling** determines when to capture experiential responses during XR exposure. Two temporal scales are addressed: (i) Fleeting Event Sampling creates Instant Assessment Items for capturing immediate, pre-cognitive responses within 3-second windows; (ii) Savouring Event Sampling develops Retrospective Assessment Items for evaluating consolidated experiences at 5-minute, 30-minute, and 24-hour intervals. **Response Congruence Formatting** provides interpretation protocols for analysing physical-virtual behavioural differences. Two analytical levels are configured: (i) Individual Difference Configuration yields the Temporal Distillation Protocol for tracking adaptation patterns within participants; (ii) Group Difference Configuration produces the Uncertainty Distillation Protocol for distinguishing system-level issues from individual preferences across participants.

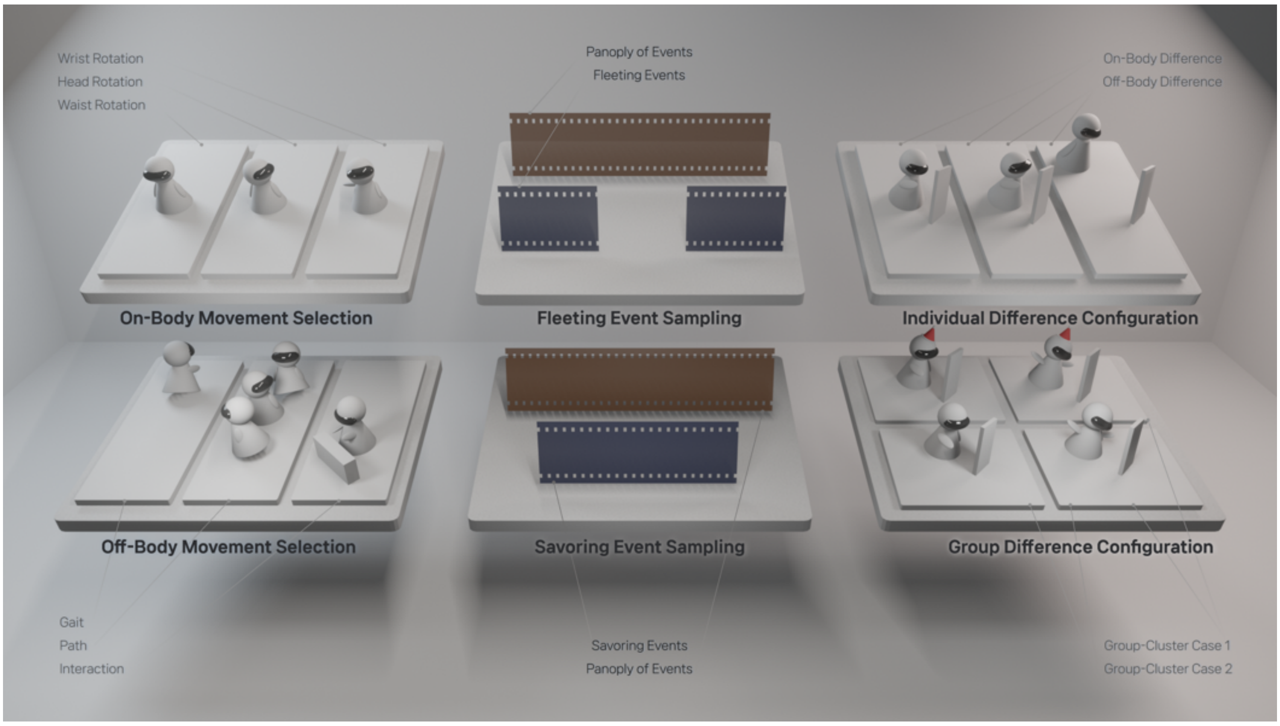

Figure 1: Overview of survey design guideline.

2.1 Target Movement Setting

On-Body Movement Selection: Understanding individual movement patterns in physical space is essential for detecting changes when virtual objects are augmented. Researchers must select appropriate on-body movements for observing the target built environment and virtual objects. Common measurements include head rotation, arm movement, hand gestures, and finger pointing. Combined movements such as head-body angle relationships and gaze-arm coordination provide richer interpretative possibilities (Lausberg & Sloetjes, 2016). Before survey deployment, baseline construction requires observing movements in non-augmented physical environments. This establishes individual normal ranges for each selected movement. Statistical normalisation and sufficient participant recruitment are necessary to exclude order effects. The Personal Pattern Baseline emerges from these individual measurements, enabling detection of meaningful deviations when users encounter virtual elements. Movement selection should align with specific spatial elements being studied.

Off-Body Movement Selection: Spatial position and movement context are inseparable from spatial experience evaluation. Researchers must define off-body movements that change the body's world coordinates. First, quantify physical constraints of the survey space: room dimensions, furniture placement, and navigable areas. Within these constraints, select relevant off-body movements. Basic patterns include navigation trajectories and gait characteristics. Additionally, consider movements involving physical environment relationships: wall distances, contact seeking behaviours, task-oriented movement (Grauman et al., 2024). Statistical normalisation remains essential. The high variance inherent in physical-virtual-body relationships necessitates self-report cross-validation. The Spatial Pattern Baseline synthesises these movement characteristics into statistical distributions. This baseline enables identification of virtual element influences on spatial behaviour. Measurement should capture both metric distances and trajectory patterns. Distinguishing meaningful changes from natural movement diversity requires careful baseline establishment before introducing virtual elements.

2.2 Affection Panoply Modelling

Fleeting Event Sampling: Movements occurring before rationalisation provide rare information for spatial experience evaluation (Slater, 2018). Researchers must identify moments capturing fleeting affection during virtual object interactions. Universal sampling points include initial virtual object recognition (0-3 seconds), boundary contact attempts, expectation violations, and sensory conflicts. Sampling strategies include event-triggered sampling (concentrated around likely fleeting affection moments) and random sampling (periodic measurement). Strategy selection depends on research objectives and virtual object characteristics. Capturing brief movements without affecting participants is crucial. Multi-angle camera recording and wrist-worn sensor indicators offer practical solutions. The critical window is the first three seconds before cognitive override occurs. These instant moments reveal authentic responses before defensive mechanisms activate. Measurement must be minimally invasive to preserve natural behaviour. Fleeting affection sampling captures raw experiential data that retrospective methods cannot access.

Savouring Event Sampling: Spatial experiences requiring time need evaluation after savouring affection develops. Dividing experiences into minimal units enables in-depth affection observation. Integration timepoints vary: 5 minutes (sensory impression), 30 minutes (emotional stabilisation), 24 hours (memory reconstruction). Researchers should align retrospective response timing with physical environment and virtual object properties. Multiple timepoint assessments capture different information layers (Hailpern et al., 2009). When constructing questionnaires, focus on specific events rather than general impressions. Recalling events for questionnaire responses requires supplementary materials that enable participants to re-evoke the original context (i.e. providing synchronised first-person and third-person video). Comparing retrospective current evaluations with original experiences reveals important differences. Response directionality—positive drift (acceptance) or negative drift (rejection)—provides measurable indicators of savouring affection changes. These temporal patterns indicate whether virtual elements integrate into experiential narratives or remain foreign intrusions. Memory reconstruction from one's movement offers insights to capture affection (Sellen & Whittaker, 2010).

2.3 Response Congruence Formatting

Individual Difference Configuration: Individual characteristics and random events create unavoidable variance in spatial experience (Höök, 2018). Multi-timepoint individual measurements should extend beyond averaging to enable deeper evaluation. The Temporal Distillation Protocol must be designed during survey planning. Rather than simple aggregation, classify repeated measures into trajectory patterns. For example, 'discomfort → neutral → comfort' and 'neutral → neutral → neutral' yield identical means but represent different experiences ('adaptation' versus 'stable neutrality'). Change magnitude becomes an additional indicator. Time series analysis and signal processing techniques can capture patterns. The protocol preserves dynamics and nuances that basic statistics miss. Individual baseline deviations beyond ±2SD indicate special events requiring investigation. The goal is not statistical generalisation but capturing physical-virtual spatial experience integration. This protocol distinguishes momentary fluctuations from persistent experiential patterns, revealing adaptation processes and habituation patterns beyond simple satisfaction scores.

Group Difference Configuration: Recognising that collected data may not represent uniform spatial experiences enhances scientific validity. Protocol design must address grouping methods and difference analysis strategies. Binary, mutually exclusive indicators for grouping are recommended. For instance, when measuring wall touching movements near virtual windows, compare groups with high x-indicator versus y-indicator frequencies. This enables the researcher's interpretation beyond statistical correlation. While continuous indicators like gait frequency can be included, they increase analytical complexity and sample requirements. Distinguishing rare events from outliers is crucial. One participant's unique movement (attempting wall passage) represents high-information events suggesting future interaction paradigms, not mere errors. Well-designed Uncertainty Distillation Protocol establishes consensus while systematically examining multi-layered uncertainties in physical-virtual integration. The protocol moves beyond majority voting to preserve unique usage patterns while identifying systematic problems requiring design intervention.

3 Proof of Concept and Discussion

To detail the Survey Design Guideline, we present a virtual window experiment examining its impact on spatial experience as an example. The study addresses two research questions: Does the virtual window affect occupant mood? Does it trigger new spatial affordances? The experiment utilises Apple Vision Pro in an 800×330 cm indoor space, where virtual windows are augmented on walls (see Fig. 2).

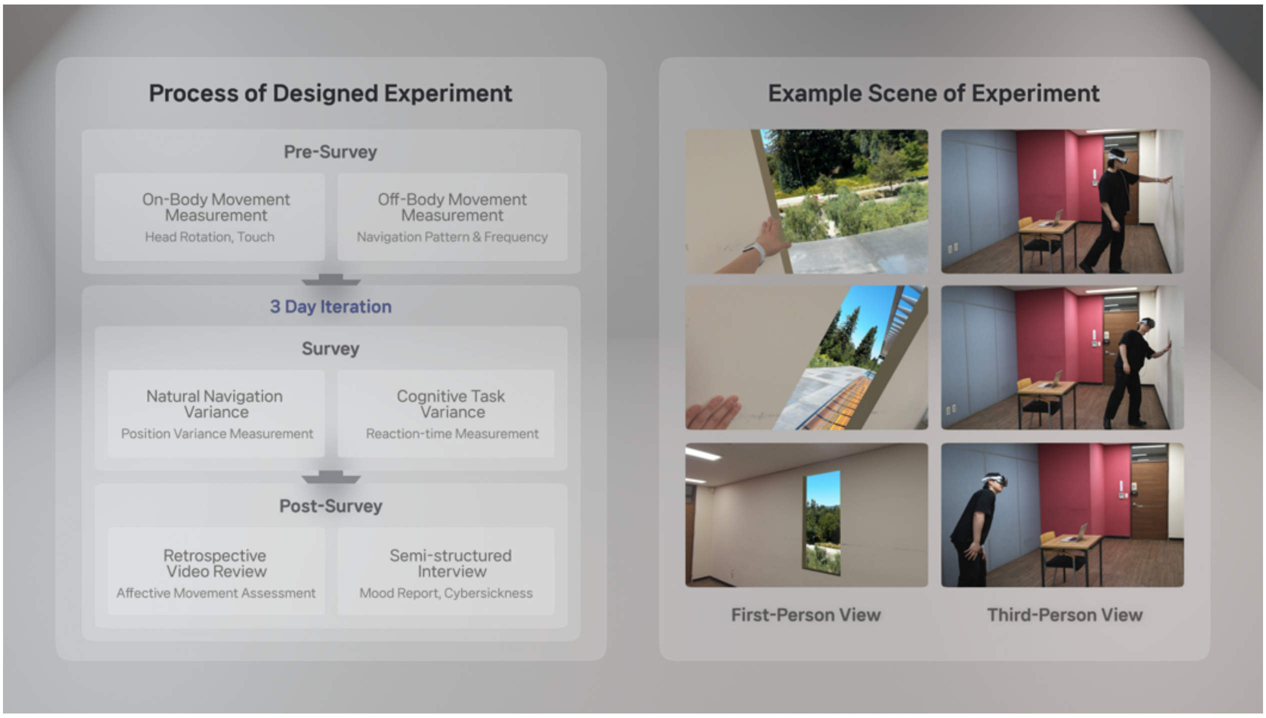

Figure 2: Overview of exemplary experiment.

3.1 Applying Target Movement Setting

For this study, we establish movement baselines through preliminary observation in the physical space without augmentation. For on-body movements, we select head rotation angle, viewing duration, hand reaching gestures, and head-body angle relationships as primary indicators. Physical window observation reveals baseline patterns: users maintain 60-80cm viewing distance, exhibit 15-30° downward gaze angles, and display 3-10 second viewing durations. These measurements form the

Personal Pattern Baseline after statistical normalisation across participants. Individual variations are documented (i.e., Taller participants show different approach angles). Deviations exceeding ±2SD from baseline indicate a significant behavioural difference.

For off-body movements, we focus on approach trajectories, stopping positions, spatial distribution, and corner avoidance radii within the room. The 800×330cm space allows a maximum 9.4m diagonal movement, with furniture creating natural navigation constraints. Physical baseline measurements show users prefer positions 2-3m from walls when no windows exist, moving primarily along room perimeters. The Spatial Pattern Baseline captures these navigation preferences through heatmap analysis and trajectory mapping after statistical normalisation. Vision Pro's inside-out tracking provides continuous position data at 90Hz. Participant movements are simultaneously recorded from two camera angles for cross-validation.

3.2 Applying Affection Panoply Modelling

Fleeting event sampling targets critical moments within the 3-second window: initial window recognition (0-3 seconds) and boundary contact attempts (within 3 seconds of approach). At these moments, the experimenter asks participants: "Right now: drawn to or away from the window?" This binary choice captures pre-cognitive responses through natural verbal collection. Event-triggered sampling activates when participants' hands enter the window's proximity zone (within 30cm), with the prompt "Touch or observe?" to identify affordance recognition. Sensory conflict moments are marked when participants show sudden movement changes or hesitation.

For savouring event sampling, we establish three assessment points. At 5 minutes, participants complete a brief verbal report about their strongest impression. At 30 minutes, after completing a reading task near the window, they rate naturalness on a 5-point scale with attention check questions embedded. At 24 hours, participants receive a follow-up questionnaire asking, "Thinking back, did the virtual window feel more or less natural than it seemed yesterday?" This temporal progression captures experience consolidation. Video recordings from both ego-centric (Vision Pro's passthrough) and third-person perspectives support retrospective assessment, synchronised and presented during the 24-hour evaluation for participants to review specific behavioural moments.

3.3 Applying Affection Panoply Modelling

Individual difference configuration employs a three-trial protocol across consecutive days. Each participant experiences the virtual window three times, allowing trajectory pattern classification. If viewing distance decreases progressively (e.g., 150cm→100cm→70cm), we code this as adaptation. If distance remains elevated despite comfort ratings, this indicates unconscious adaptation or ignorance. Patterns deviating beyond ±2SD from the Personal Pattern Baseline receive additional analysis. The Temporal Distillation Protocol preserves these patterns rather than averaging them, recognising that Day 3 responses carry more weight than Day 1 reactions. Attention checks validate response consistency across trials. We conducted this analysis following our experimental principles and visualised the results by individual and temporal conditions (see Fig. 3)

Group difference configuration separates participants by interaction patterns using binary, mutually exclusive indicators. Binary grouping distinguishes ‘approachers’ (final distance <100cm) from "observers" (maintaining >100cm). When most participants maintain distances exceeding physical window baselines (60-80cm) despite reporting satisfaction, this systematic divergence indicates design issues rather than individual preferences. The specific threshold emerges from data distribution analysis rather than predetermined values. Rare events receive special attention—if one participant attempts to "open" the virtual window while others don't, this represents high-information events suggesting latent affordances. The Uncertainty Distillation Protocol identifies both convergent problems requiring immediate fixes and divergent behaviours suggesting future design opportunities.

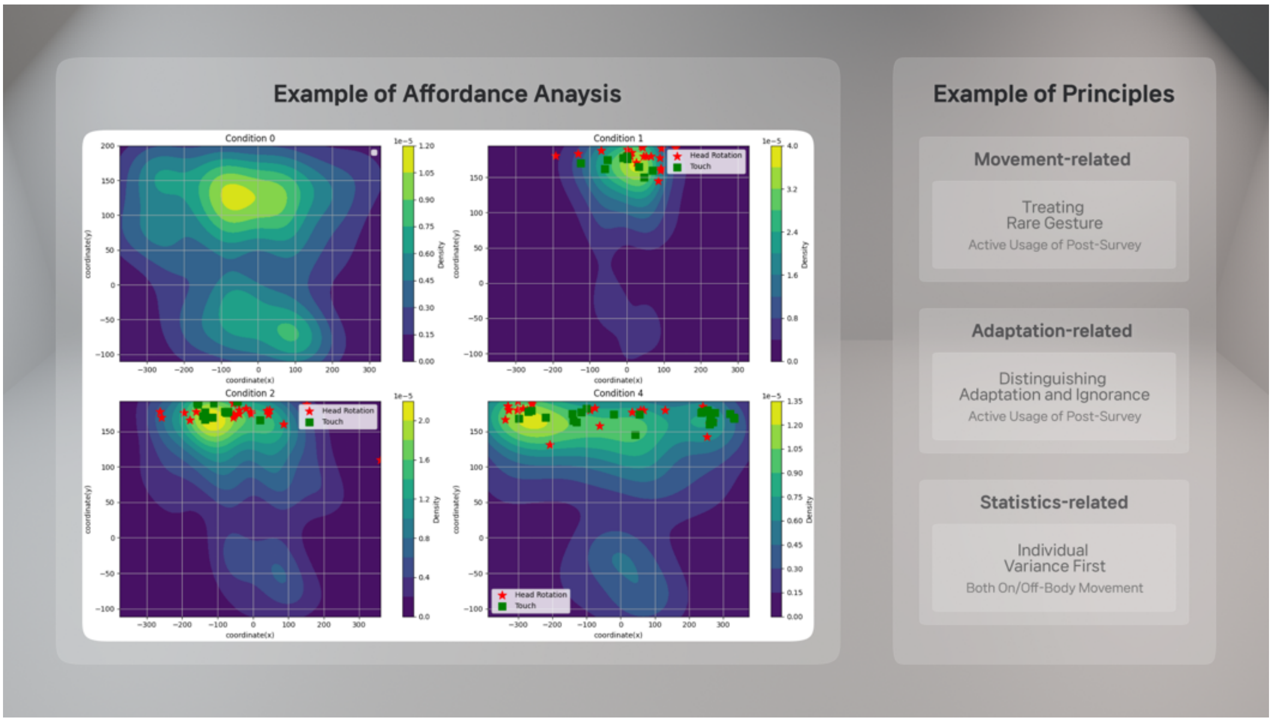

Figure 3: Example of data analysis and principle for analysis.

3.4 Gesture Analysis Direction for Indoor Spatial Experiment Evaluation

Successfully integrated virtual windows present both clear behavioural convergence for intention and unique differences with physical window patterns by Day 3. Mood effects manifest through increased time spent near windows and approach frequencies within ±2SD of baseline measurements. New affordances emerge through unexpected interactions—reaching, opening attempts, or using the window during work (cognitive task). Failure patterns include distance increase after Day 1, reduced viewing duration, and absence of natural behaviours like seeing during break due to sparse changes of scene. These behavioural indicators, combined with temporal assessment data, provide a comprehensive evaluation of whether virtual windows achieve genuine spatial integration or remain foreign overlays in the physical environment. After the experiment, retrospective video analysis uses both egocentric and third-person perspectives. These materials support multidimensional analysis of spatial experience through gesture-focused interviews (Subramanian et al., 2021).

4 Conclusion and Future Work

This paper presented SEEMEE, a survey design guideline capturing behavioural responses in XR spatial experience evaluation through behavioural congruence measurement. The three-stage guideline—Target Movement Setting, Affection Panoply Modelling, and Response Congruence Formatting—was detailed through an exemplary virtual window experiment demonstrating practical application. Future work will apply this guideline to diverse XR applications, documenting context-specific considerations across architectural, retail, and workplace environments. This systematic gesture-based analysis will contribute to a more authentic evaluation of spatial experience.

This paper presented SEEMEE, a survey design guideline capturing behavioural responses in XR spatial experience evaluation through behavioural congruence measurement. The three-stage guideline—Target Movement Setting, Affection Panoply Modelling, and Response Congruence Formatting—was detailed through an exemplary virtual window experiment demonstrating practical application. Future work will apply this guideline to diverse XR applications, documenting context-specific considerations

across architectural, retail, and workplace environments. This systematic gesture-based analysis will contribute to a more authentic evaluation of spatial experience.

Funding Acknowledgement
This work was supported by STEAM R&D Project, NRF, Korea (RS-2024-00454458).

Ethical Approval Declaration
The study was conducted in accordance with established standards for research integrity and ethics.

Data Availability Statement
Data can be made available upon request to the corresponding author.

Conflicts of Interest
The authors declare no conflict of interest.

References

Brooke, J. (1996). SUS: A “quick and dirty” usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–194). Taylor & Francis.

Grauman, K., Westbury, A., Torresani, L., Kitani, K., Malik, J., Afouras, T., ... Grabler, F. (2024). Ego-Exo4D: Understanding skilled human activity from first- and third-person perspectives. *arXiv preprint*. https://doi.org/10.48550/arXiv.2311.18259

Hailpern, J., Karahalios, K., Halle, J., Dethorne, L., & Coletto, M.-K. (2009). A3: HCI coding guideline for research using video annotation to assess behavior of nonverbal subjects with computer-based intervention. *ACM Transactions on Accessible Computing*, *2*(2), 1–29. https://doi.org/10.1145/1530064.1530066

Halbig, A., & Latoschik, M. E. (2021). A systematic review of physiological measurements, factors, methods, and applications in virtual reality. *Frontiers in Virtual Reality*, *2*, 694567. https://doi.org/10.3389/frvir.2021.694567

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). North-Holland. https://doi.org/10.1016/S0166-4115(08)62386-9

Höök, K. (2018). *Designing with the body: Somaesthetic interaction design*. MIT Press.

Lausberg, H., & Sloetjes, H. (2016). The revised NEUROGES–ELAN system: An objective and reliable interdisciplinary analysis tool for nonverbal behavior and gesture. *Behavior Research Methods*, *48*(3), 973–993. https://doi.org/10.3758/s13428-015-0622-z

Preiser, W. F. E., White, E., & Rabinowitz, H. (2015). *Post-occupancy evaluation* (Routledge Revivals). Routledge. https://doi.org/10.4324/9781315713519

Sellen, A. J., & Whittaker, S. (2010). Beyond total capture: A constructive critique of lifelogging. *Communications of the ACM*, *53*(5), 70–77. https://doi.org/10.1145/1735223.1735243

Slater, M. (2018). Immersion and the illusion of presence in virtual reality. *British Journal of Psychology*, *109*(3), 431–433. https://doi.org/10.1111/bjop.12305

Subramanian, K., Maas, J., Borchers, J., & Hollan, J. (2021). From detectables to inspectables: Understanding qualitative analysis of audiovisual data. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–10). ACM. https://doi.org/10.1145/3411764.3445458