Enhancing mHealth App Onboarding Using a Multimodal Mixed Reality Avatar

Stefan Resch
Frankfurt UAS
Frankfurt am Main, Germany
University of Cadiz
Cadiz, Spain
stefan.resch@fra-uas.de

Yunus Söyleyici
Frankfurt UAS
Frankfurt am Main, Germany
yunus.soeyleyici@stud.fra-uas.de

Valentin Schwind
Stuttgart Media University
Stuttgart, Germany
schwind@hdm-stuttgart.de

Diana Völz
Frankfurt UAS
Frankfurt am Main, Germany
voelz@fra-uas.de

Daniel Sanchez-Morillo
University of Cadiz
Cadiz, Spain
daniel.morillo@uca.es

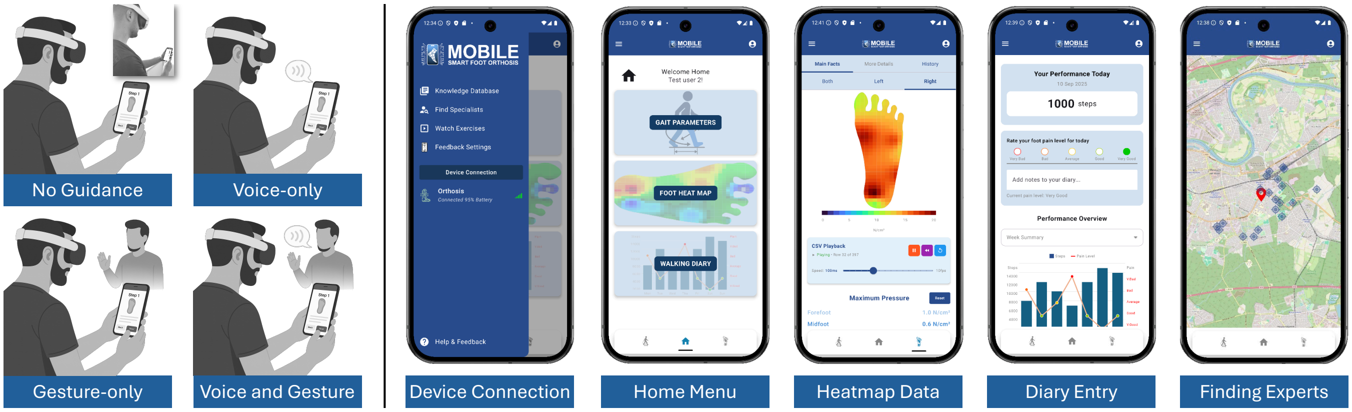


Figure 1: Schematic overview of the MR onboarding setup: four conditions (left) and representative app functions (right).

Abstract

Navigating novel smartphone applications is often complex and unintuitive during the initial onboarding phase. Missing or ineffective guidance can cause misuse, errors, and abandonment, which is particularly critical in domains such as mobile health (mHealth) apps. While previous research has shown potential for onboarding with conversational avatars in mixed reality (MR), the impact of avatars combining gestures and voice for app tutorial guidance remains unknown. To address this gap, we conducted a within-subject study with 24 participants using a smart insole mHealth app as a case example. Participants completed tasks in MR under four conditions varying the presence of avatar voice and animated pointing gestures. Our findings show that tutorial guidance improved with voice and gesture compared to no avatar support, enhancing usability and learning experience while also highlighting remaining challenges. We contribute design implications for avatar-supported guidance in mHealth and broader human-computer interaction (HCI) contexts.

CHI EA '26, Barcelona, Spain

ACM ISBN 979-8-4007-2281-3/26/04
https://doi.org/10.1145/3772363.3799298

CCS Concepts

- **Human-centered computing → Human computer interaction (HCI)**; **Empirical studies in HCI**; **Mixed / augmented reality**; **Smartphones**.

Keywords

Avatar, Mixed Reality, Smartphone, Onboarding, Mobile Health App, Learning Experience, Usability

ACM Reference Format:
Stefan Resch, Yunus Söyleyici, Valentin Schwind, Diana Völz, and Daniel Sanchez-Morillo. 2026. Enhancing mHealth App Onboarding Using a Multimodal Mixed Reality Avatar. In *Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems (CHI EA '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3772363.3799298

1 Introduction and Background

The interaction and navigation of novel smartphone applications remain challenging, particularly during first-time use. This issue is especially critical in mHealth applications, where navigating health-related data can be complex, and correct interaction as well as accurate data interpretation are essential for effective use and long-term adoption [18, 34]. Prior work shows that missing or ineffective guidance often leads to misuse, errors, and abandonment, as usability and user satisfaction are strongly linked to successful use [17, 29], highlighting the need for structured onboarding and learning support. In the rapidly growing field of mHealth applications [15, 25], smart insole systems show potential for foot health monitoring and rehabilitation via smartphone apps [10, 14, 32]. However, prior work reports usability challenges, particularly during initial interaction and self-guided learning [27, 28]. Consequently, effective onboarding strategies are a key requirement for successful user–device interaction.

As shown by prior work, targeted onboarding in new applications can support first-time users through tutorials, increasing engagement and learnability [8]. Such onboarding strategies have been tested in various domains, including employee training and consumer apps [12, 23, 30, 40, 41]. For example, Froehlich et al. [12] demonstrated that first-time onboarding in a cryptocurrency app improved usability and informed instructional design. In line with design guidelines proposed by Kascak et al. [19, 20], Ruzic et al. [36] highlight that combining pictorial and verbal presentations helps maintain users' information awareness, an effect also observed for visual and auditory notifications in mHealth apps [45]. However, it remains unclear how these modalities contribute to effective onboarding in the context of mHealth apps.

For this reason, MR offers new opportunities for interactive onboarding through embodied and multimodal guidance [7, 44], and has also been explored in gamified healthcare apps using avatars [13]. Virtual avatar–based approaches have also been explored beyond healthcare, for example in machine-task tutoring, demonstrating their potential for instructional support [6, 16]. Prior work has shown that conversational avatars in MR can support learning, for example in programming education, where virtual avatars enabled natural interaction and improved user experience [22, 24]. Furthermore, combining gesture and voice modalities can enhance user experience in virtual teaching [11] and guidance [46] environments, while visual pointing cues help guide attention to relevant content [31]. Building on this, Weichbroth [43] emphasized that future mobile app usability research should investigate multimodal MR interactions using visual and audio content.

However, while avatar-based learning in MR shows promise, it remains unclear how voice and gesture guidance affect usability and tutorial experience during onboarding in mHealth applications. To address this gap, the following research question arises: *How does avatar-based voice and gesture guidance in MR influence usability, learning experience, and perceived tutorial guidance during onboarding of a smart insole application?* Therefore, we conducted a within-subject study with 24 participants using a smart insole mHealth app, investigating the effect of an MR avatar combining visual gestures and audible voice guidance on usability and learning/tutorial experiences during app onboarding.

Our findings provide initial insights to inform the future design of onboarding systems for mHealth apps and motivate follow-up investigations. We contribute: (1) empirical insights into multimodal, avatar-based onboarding in mHealth, showing how gesture and voice influence user interactions; and (2) design implications for avatar-supported mHealth onboarding, emphasizing adaptive explanations and interactive steps to enhance learning and usability.

2 Method

We conducted a mixed-method user study with 24 participants to examine the effects of avatar-based voice explanations and pointing instructions on mHealth app onboarding in MR. The study focused on usability, learning experience, and perceived tutorial guidance during first-time app use. Here, learning experience refers to users' immediate, subjective instructional experience during the tutorial, reflecting their understanding and confidence. The study focused on initial tutorial experiences via avatar-based guidance, not long-term outcomes, as app content and tasks were identical across trials. A smart insole mHealth app was chosen, as such apps require learning both interaction flows and sensor-based data interpretation, creating cognitively demanding onboarding scenarios [34].

We used a 2×2 within-subjects design with the independent variables AVATAR VOICE (*audible* vs. *non-audible*) and AVATAR GESTURE (*visible* vs. *non-visible*). In the baseline condition (no avatar guidance), participants navigated the app based solely on the written task description, without avatar-based audio or gesture guidance. All conditions used identical tasks to isolate the effects of voice, gesture, and their combination. Condition order was counterbalanced using a balanced Latin square to avoid sequencing effects [2].

2.1 Stimuli

The study used a smart insole mHealth app co-designed with users and experts [35] and previously tested for usability [33, 34], ensuring comparability across studies. Key functions included monitoring gait, a foot heatmap, a walking diary, and finding specialists. These functions formed the basis of the study tasks, see Figure 1.

A life-sized, full-body female avatar was used, reflecting common practice in related studies for comparability [47] and to increase acceptance of the avatar's hands [39]. We used the "Kate" model as MR avatar with the “Pointing” animation in its seated version from Mixamo. The avatar held a virtual smartphone in its left hand, matching the real device dimensions and displaying task videos recorded on the actual device. During the tutorial, the avatar pointed with its right hand to relevant interface elements (e.g., buttons, menu icons, text fields), giving the impression of real-time interaction with the device (see Figure 2). The animation and video looped, maintaining the avatar's interactive appearance until participants performed the corresponding action on the physical device. In the voice-and-gesture condition, stepwise spoken instructions were synchronized with the pointing gestures using Luvvoice [1] (“Aria” female model).

2.2 Apparatus

The mobile app was developed in Android Studio (v. 2024.2.1) and deployed as an Android package (APK) on a Samsung Galaxy S21 5G (Android 15, SDK v. 35). Pre-recorded demo heatmap data were visualized as gait trials, and avatar display videos were recorded on the device for MR synchronization. The Meta Quest 3 served as head-mounted display (HMD) with video passthrough for MR. Development employed Unity engine (v. 6000.0.45f1) and the Meta XR All-in-One SDK [26] (v. 72.0.0). Interviews were audio-recorded with a Logitech C922 Pro HD Stream webcam and Open Broadcaster Software. All software ran on a Windows 11 Pro (XMG Fusion 15, i7-11800H, 8-core 2.3 GHz, RTX 3070, 16 GB RAM).

---

[1] https://luvvoice.com/

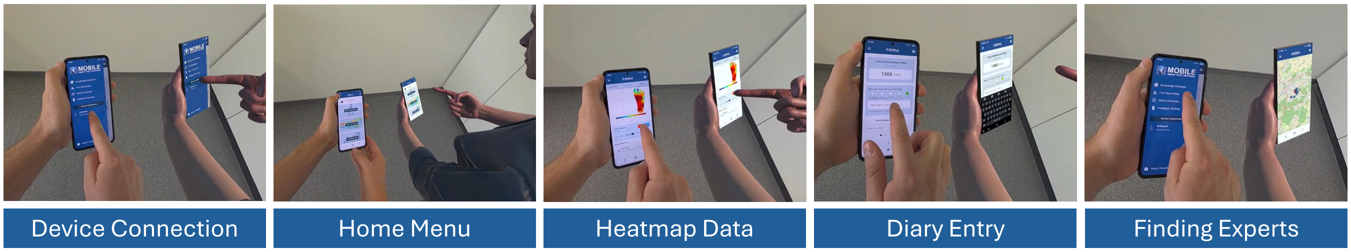


Figure 2: Five screenshots from the MR condition with avatar gestures, illustrating different task steps: connecting the device, navigating the home menu, explaining the heatmap using pre-recorded data, entering a diary entry, and locating an expert.

2.3 Procedure

After the study procedure was explained, participants provided informed consent and completed a demographic questionnaire. They were then equipped with the HMD and completed a short reading test using video passthrough to ensure sufficient task visibility. Before starting, they were instructed to focus on tutorial guidance rather than speed, as the usability and learning experience assessments targeted the tutorial type and guidance rather than the app itself. Depending on the condition, participants were guided through the app by the avatar (voice and/or gestures), which was scaled and positioned consistently next to them, or explored the app independently. After each condition, a semi-structured interview was conducted and audio recorded, followed by the System Usability Scale (SUS), Learning Experience Questionnaire (LEQ), and Tutorial Experience Questionnaire (TEQ) questionnaires. The procedure was repeated for all conditions, with a final debriefing, and lasted approximately 90 minutes per participant.

2.4 Tasks

The app tutorial was structured around predefined tasks to provide a representative onboarding procedure covering the app's core functions. Tasks were selected to reflect typical user goals and involve comparable interaction steps across all conditions, ensuring that differences in performance or perception can be attributed to the tutorial approach rather than the tasks themselves. They included navigating the home menu, connecting the smart insole, testing basic functions, interpreting heatmap data, adding entries to the walking diary, and finding specialists. Participants were also asked to search for specific functions, simulating daily situations requiring navigation of the user interface (UI).

The simplified task descriptions are provided below and were read before the first condition to ensure task understanding. For the control condition without guidance, the description was read a second time to ensure a fair comparison.

- Task 1: Connect the smartphone app to the smart insole. Assume that you do not need to leave the app or interact with anything outside of it to complete this task.
- Task 2: Access your recorded foot pressure data, view it as a heatmap, and explore the playback function.
- Task 3: Open your walking diary, record your current pain level, and add a short note.
- Task 4: Use the integrated map to find a specialist in your area and display detailed information about them.

2.5 Measures and Data Analysis

Quantitative subjective data were analyzed using descriptive and inferential statistics, including repeated measures (RM) Analysis of Variance (ANOVA), in R with the rstatix package [21]. The SUS [4] was used to assess tutorial usability and has been applied in prior onboarding evaluations [12, 49]. Following the standard scoring procedure, item ratings were transformed to yield scores ranging from 0 to 100 and interpreted using established usability benchmarks [1, 4]. The LEQ [48] was used to assess the perceived learning experience of the avatar modalities. It comprises 19 items across four factors (contextuality, effects, design features, and explorative nature) rated on a five-point Likert scale. The TEQ [38] assessed tutorial experience in digital environments. It comprises eight items across two subscales (structural clarity and transparency) rated on a seven-point Likert scale from -3 (*strongly disagree*) to 3 (*strongly agree*), with 0 as neutral. Overall LEQ/TEQ and subscale scores were computed as the mean per participant and condition.

Qualitative data were collected via semi-structured interviews on participants' perceptions of voice and gesture modalities, allowing flexible follow-up questions. In total, 96 interviews were conducted (182 min of audio, $M = 113$ s, $SD = 60$ s) and transcribed using Buzz [2]. Transcripts were checked and corrected by one researcher, with a second verifying completeness. An inductive thematic analysis [3] was performed independently by two researchers in an iterative process to identify key categories and themes across the dataset. First, the transcribed dataset was coded sentence by sentence (open coding). Following this, axial coding was applied to cluster categories by identifying relationships, and selective coding was used to refine the overarching thematic structure. Differences between coders were resolved through iterative discussion and comparison of coding decisions until consensus was reached, ensuring the final themes accurately reflected the data.

2.6 Participants

The study received ethical clearance from the German Society for Nursing Science (No. 23-027). We recruited 24 participants (13 male, 11 female), aged 21–61 ($M = 32.83$, $SD = 12.63$), via personal contacts and institutional mailing lists. The sample size follows the standards for HCI studies [5]. Participants came from five nationalities (Europe, North America, and Asia) and had diverse educational backgrounds and occupations. All reported prior experience with health-related devices and apps, but none had previously used a

---

[2] https://github.com/chidiwilliams/buzz.git

smart insole app. Familiarity with mHealth apps averaged 2.71 ($SD = 1.20$) on a five-point Likert scale, indicating a moderate level of familiarity. Half of the participants ($n = 12$) wore glasses. No participants dropped out and all were included in the data analysis.

3 Results

3.1 Quantitative Results

3.1.1 System Usability Scale (SUS). Shapiro–Wilk tests confirmed normality for the overall SUS score ($p > .065$). A two-way RM ANOVA revealed significant main effects of GESTURE ($F(1, 23) = 11.05$, $p = .003$, $\eta_p^2 = .32$) and VOICE ($F(1, 23) = 6.46$, $p = .018$, $\eta_p^2 = .22$), as well as a significant interaction ($F(1, 23) = 6.78$, $p = .016$, $\eta_p^2 = .23$), all with large effect sizes. Post-hoc tests with Bonferroni correction (Kenward–Roger) showed that VOICE was significant only when the AVATAR was *non-visible* ($p < .001$), and AVATAR was significant only when VOICE was *non-audible* ($p < .001$), see Figure 3a. Analysis of SUS subscales confirmed this pattern, showing significant effects of GESTURE and VOICE across six dimensions (all $p < .038$): frequency of use, ease of use, inconsistency, learnability, confidence, and prior learning (see Appendix A.1: Table 1 and Figure 5).

3.1.2 Learning Experience Questionnaire (LEQ). Shapiro–Wilk tests on LEQ ratings confirmed a normal distribution (all $p > .523$). A two-way RM ANOVA revealed significant main effects of GESTURE ($F(1, 23) = 4.57$, $p = .043$, $\eta_p^2 = .166$) and VOICE ($F(1, 23) = 14.11$, $p = .001$, $\eta_p^2 = .380$), and a significant interaction ($F(1, 23) = 10.34$, $p = .004$, $\eta_p^2 = .310$), all with large effect sizes. Post-hoc Bonferroni tests (Kenward–Roger) showed that *audible* voice > *non-audible* in the *non-visible* gesture condition ($t(69) = 3.91$, $p < .001$), and *visible* > *non-visible* in the *non-audible* voice condition ($t(69) = -3.30$, $p = .002$), see Figure 3b. Furthermore, RM ANOVA on subscales revealed significant effects for three dimensions (contextuality, effects, design features, all $p < .011$), but none for explorative nature of learning experience, see Appendix A.2 (Table 2 and Figure 6).

3.1.3 Tutorial Experience Questionnaire (TEQ). Shapiro–Wilk tests indicated non-normality of the overall TEQ score. An aligned rank transform (ART) RM ANOVA revealed significant main effects of GESTURE ($F(1, 69) = 6.01$, $p = .002$) and VOICE ($F(1, 69) = 24.04$, $p < .001$), but no interaction ($F(1, 69) = 0.92$, $p = .341$). Post-hoc Bonferroni tests showed higher ratings for *visible* > *non-visible* ($p = .021$) and *audible* > *non-audible* ($p < .001$), see Figure 3c. For the two subscales, ART RM ANOVA revealed main effects of GESTURE ($F(1, 69) = 6.30$, $p = .014$), VOICE ($F(1, 69) = 21.28$, $p < .001$), and an interaction ($F(1, 69) = 4.71$, $p = .003$) for structural clarity, as well as a main effect of VOICE ($F(1, 69) = 14.70$, $p < .001$) for transparency. Post-hoc tests indicated that for structural clarity, *visible* avatars scored higher when VOICE was *non-audible* ($p = .002$), and *audible* voice increased scores when AVATAR was *non-visible* ($p < .001$); for transparency, higher ratings were observed for *visible* > *non-visible* ($p = .009$) and *audible* > *non-audible* ($p = .001$). The subscale results are documented in Appendix A.3 (Figure 7).

3.1.4 Preferred Conditions (Likert Ratings). After the experiment, participants rated their favored combination of GESTURE and VOICE ($N = 15$), followed by VOICE-only ($N = 7$), and NO GUIDANCE ($N = 2$). Participants indicated that a talking avatar in MR could be helpful for explaining and interpreting health data (22 out of 24). Overall satisfaction with the MR app guidance was high ($M = 5.83$, $SD = 1.20$ on a 7-point Likert scale).

3.2 Qualitative Results: Semi-Structured Interview

Inductive thematic analysis revealed three key themes.

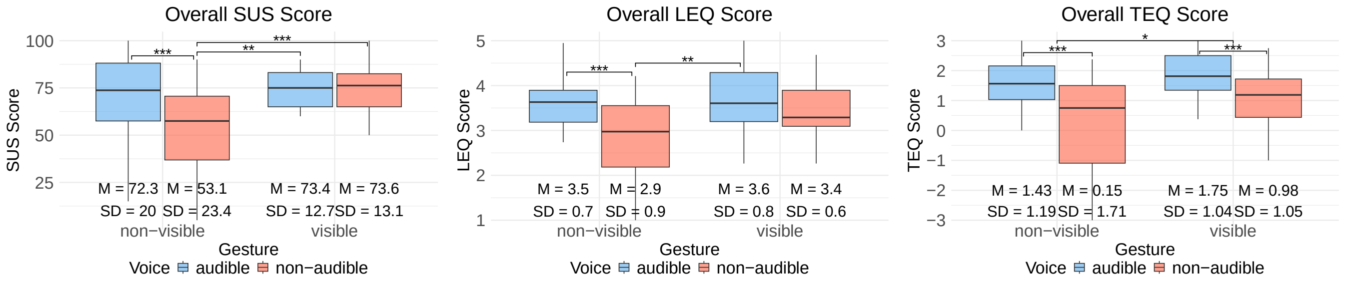
1. Engagement in Feedback: The baseline was described as "convenient" (P8) and suitable for self-directed learning (P19); voice guidance as "clear and understandable" (P11) and "effective for learning new information" (P21); and the combined voice–gesture condition as "interactive" (P17), "easy to use" (P13), and "confidence-enhancing" (P5), providing "clear instructions and visual cues" (P9).

2. Guidance Challenges: No guidance was criticized as "completely overloaded" (P24) and "difficult to navigate" (P12, P19); avatar-only led to "clicking without thinking" (P11) and limited learning (P11, P21); voice-only was described as "strenuous or confusing" (P9); and the combined condition as "distracting" (P19).

3. Reflections & Suggestions: Despite an overall positive experience, the avatar was mainly perceived through its hands: "I didn't really pay attention to him as a person, only to the hand" (P10); "I'd prefer a larger display without an avatar and only a hand" (P21). Furthermore, avatar placement adjustments could also improve the interaction, as stated by P21: "'I tried standing inside the avatar … and then quickly repeated that on my own display." Device-dependent issues were mentioned, e.g., that MR vision "should be improved" (P4) and made "clearer" (P23).

4 Discussion

In this study, we investigated how MR avatar voice and gesture modalities affect usability and tutorial experience during onboarding of a smart insole app. Both modalities, individually and combined, improved SUS, LEQ, and TEQ scores and increased overall satisfaction with MR-based guidance. The no-guidance condition was consistently least preferred, while gesture-only was rated lower than voice-only or combined modalities, likely due to its limited contextual information. SUS results indicate that lack of guidance increases perceived complexity, inconsistency, and reliance on prior knowledge, aligning with prior work showing barriers in self-guided mHealth onboarding [28]. Moreover, avatar modalities enhanced perceived ease of use, learnability, and confidence, supporting frequent app use. While multimodal guidance increased tutorial transparency, adding a second modality did not further improve clarity, suggesting that guidance presence matters more than modality number. Notably, the exploratory aspect of learning was not hindered by guided conditions. Qualitative findings show participants primarily attended to the avatar's hands and gestures, rather than the full body, highlighting the importance of functional embodiment. Furthermore, onboarding should also align with task type: gestures support procedural tasks (menu navigation, device connection, locating specialists, or diary entry) by improving spatial orientation and replication, while voice is more effective for interpretative tasks (gait metrics, pressure heatmaps, diary graphs, or health information) due to its narrative clarity. Combined modalities were perceived as more demanding, suggesting



(a) System Usability Scale **(b) Learning Experience Questionnaire** **(c) Tutorial Experience Questionnaire**

Figure 3: Boxplots of overall questionnaire ratings: (a) SUS, (b) LEQ, and (c) TEQ, including descriptive statistics.

selective use when understanding affects health-related decisions. Taken together, these findings indicate MR avatar onboarding is most effective when modalities match mHealth tasks.

Design Implications: We propose the following context-specific design implications: *(1) Modality contextualization:* Guidance in MR mHealth apps should align with task demands. Procedural steps benefit from gestures, interpretative steps from voice. Multimodal support may increase cognitive demand and be toggled by user experience or health literacy. *(2) Avatar embodiment:* Our results indicate that participants primarily attended to hands and gestures, suggesting that full-body embodiment may be more relevant for other mHealth tasks. Aligning gesture orientation with the user's viewpoint and providing egocentric augmentations (e.g., rendering virtual hands in the user's field of view) can improve clarity and reduce misinterpretation or distraction. *(3) Interactive guidance:* mHealth guidance should let users actively follow and repeat interactive steps, reinforcing health data understanding, supporting confidence and active engagement in self-monitoring. Repeating critical task steps can support procedural memory [9], particularly when learning to interpret gait and pressure parameters. *(4) Device-centered augmentation:* Several participants reported readability issues with the virtual smartphone model, highlighting the need for improved visibility. MR smartphone onboarding should therefore consider task-specific augmentations, such as enlarged or adaptive displays, to enhance legibility. The choice of HMD also substantially affects visibility and user comfort. While the Meta Quest 3 provides moderate passthrough quality [42], optical see-through augmented reality (AR) headsets can mitigate visibility limitations and should be considered when designing avatar-based onboarding.

Limitations and Future Work: Our study is limited by the selected sample and a single mHealth app, constraining generalizability. The selected HMD constituted a major limitation, as reduced screen visibility may have influenced task performance. Therefore, future work should examine optical see-through displays to improve visual perception. Moreover, alternative avatar perspectives (e.g., first-person perspective) should be investigated, as suggested by prior work [37], to further optimize guidance effectiveness. Follow-up studies should replicate and extend these results with a broader range of mobile apps, including commercial mHealth systems, to validate our findings and inform effective onboarding design based on user preferences. Overall, our work provides a foundation for follow-up research on MR avatar-based guidance in mobile apps.

Acknowledgments

This research was funded by the Hessian Ministry of Science and Art - HMWK, Germany (FL1, Mittelbau) and supported by the Frankfurt University of Applied Sciences and the University of Cadiz.

References

[1] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies* 4, 3 (2009), 114–123.
[2] James V Bradley. 1958. Complete counterbalancing of immediate sequential effects in a Latin square design. *J. Amer. Statist. Assoc.* 53, 282 (1958), 525–528.
[3] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
[4] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.
[5] Kelly Caine. 2016. Local Standards for Sample Size at CHI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 981–992. doi:10.1145/2858036.2858498
[6] Yuanzhi Cao, Xun Qian, Tianyi Wang, Rachel Lee, Ke Huo, and Karthik Ramani. 2020. An Exploratory Study of Augmented Reality Presence for Tutoring Machine Tasks. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3313831.3376688
[7] Edwige Chauvergne, Martin Hachet, and Arnaud Prouzeau. 2023. User Onboarding in Virtual Reality: An Investigation of Current Practices. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 711, 15 pages. doi:10.1145/3544548.3581211
[8] Antonio Chiappetta. 2019. Designing effective user onboarding experiences for mobile applications.
[9] Michael D Cohen and Paul Bacdayan. 1994. Organizational routines are stored as procedural memory: Evidence from a laboratory study. *Organization science* 5, 4 (1994), 554–568.
[10] Luigi D'arco, Graham Mccalmont, Haiying Wang, and Huiru Zheng. 2024. Application of smart insoles for recognition of activities of daily living: A systematic review. *ACM Transactions on Computing for Healthcare* 5, 1 (2024), 1–34.
[11] Ke Fang and Jing Wang. 2024. Interactive Design With Gesture and Voice Recognition in Virtual Teaching Environments. *IEEE Access* 12 (2024), 4213–4224. doi:10.1109/ACCESS.2023.3348846
[12] Michael Froehlich, Charlotte Kobiella, Albrecht Schmidt, and Florian Alt. 2021. Is it Better With Onboarding? Improving First-Time Cryptocurrency App Experiences. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference* (Virtual Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 78–89. doi:10.1145/3461778.3462047
[13] Yu Fu, Yan Hu, and Veronica Sundstedt. 2022. A Systematic Literature Review of Virtual, Augmented, and Mixed Reality Game Applications in Healthcare. *ACM Trans. Comput. Healthcare* 3, 2, Article 22 (March 2022), 27 pages. doi:10.1145/3472303
[14] Nagaraj Hegde, Edward Melanson, and Edward Sazonov. 2016. Development of a real time activity monitoring Android application utilizing SmartStep. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, Orlando, FL, USA, 1886–1889. doi:10.1109/EMBC.2016.7591089

[15] Martin Hensher, Paul Cooper, Sithara Wanni Arachchige Dona, Mary Rose Angeles, Dieu Nguyen, Natalie Heynsbergh, Mary Lou Chatterton, and Anna Peeters. 2021. Scoping review: development and assessment of evaluation frameworks of mobile health apps for recommendations to consumers. *Journal of the American Medical Informatics Association* 28, 6 (2021), 1318–1329.

[16] Gaoping Huang, Xun Qian, Tianyi Wang, Fagun Patel, Maitreya Sreeram, Yuanzhi Cao, Karthik Ramani, and Alexander J. Quinn. 2021. AdapTutAR: An Adaptive Tutoring System for Machine Tasks in Augmented Reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 417, 15 pages. doi:10.1145/3411764.3445283

[17] Muhammad Nazrul Islam, Md Mahboob Karim, Toki Tahmid Inan, and AKM Najmul Islam. 2020. Investigating usability of mobile health applications in Bangladesh. *BMC medical informatics and decision making* 20, 1 (2020), 19.

[18] Seungwan Jin, Bogoan Kim, and Kyungsik Han. 2025. "I Don't Know Why I Should Use This App": Holistic Analysis on User Engagement Challenges in Mobile Mental Health. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1185, 23 pages. doi:10.1145/3706598.3713732

[19] Ljilja Ruzic Kascak, Seunghyun "Tina" Lee, Elaine Yilin Liu, and Jon A. Sanford. 2015. Universal Design (UD) Guidelines for Interactive Mobile Voting Interfaces for Older Adults. In *Universal Access in Human-Computer Interaction. Access to the Human Environment and Culture*, Margherita Antona and Constantine Stephanidis (Eds.). Springer International Publishing, Cham, 215–225.

[20] Ljilja Ruzic Kascak, Claudia B. Rébola, and Jon A. Sanford. 2014. Integrating Universal Design (UD) Principles and Mobile Design Guidelines to Improve Design of Mobile Health Applications for Older Adults. In *Proceedings of the 2014 IEEE International Conference on Healthcare Informatics (ICHI '14)*. IEEE Computer Society, USA, 343–348. doi:10.1109/ICHI.2014.54

[21] Alboukadel Kassambara. 2023. *rstatix: Pipe-Friendly Framework for Basic Statistical Tests*. The R Foundation. https://rpkgs.datanovia.com/rstatix/ R package version 0.7.2.

[22] Joonyoung Kim, Sudeep Agarwal, Kristina Marotta, Siwei Li, Jonathan Leo, and Duen Horng Chau. 2019. Mixed Reality for Learning Programming. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children* (Boise, ID, USA) *(IDC '19)*. Association for Computing Machinery, New York, NY, USA, 574–579. doi:10.1145/3311927.3325335

[23] Howard J Klein, Beth Polin, and Kyra Leigh Sutton. 2015. Specific onboarding practices for the socialization of new employees. *International Journal of Selection and Assessment* 23, 3 (2015), 263–283.

[24] Gilda Manfredi, Ugo Erra, and Gabriele Gilio. 2023. A Mixed Reality Approach for Innovative Pair Programming Education with a Conversational AI Virtual Avatar. In *Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering* (Oulu, Finland) *(EASE '23)*. Association for Computing Machinery, New York, NY, USA, 450–454. doi:10.1145/3593434.3593952

[25] Inocencio Maramba, Arunangsu Chatterjee, and Craig Newman. 2019. Methods of usability testing in the development of eHealth applications: a scoping review. *International journal of medical informatics* 126 (2019), 95–104.

[26] Meta. 2025. Meta XR All-in-One SDK. Unity Asset Store, https://assetstore.unity.com/packages/tools/integration/meta-xr-all-in-one-sdk-269657. Accessed: January 19, 2026.

[27] Khadijeh Moulaei and Abbas Sheikhtaheri. 2025. Usability evaluation of wearable technology: A pilot study on a smart diabetic shoe for foot care. *Clinical eHealth* 8 (2025), 94–102. doi:10.1016/j.ceh.2025.04.004

[28] Rajna Ogrin, Rekha Viswanathan, Tracy Aylen, Fiona Wallace, Janine Scott, and Dinesh Kumar. 2018. Co-design of an evidence-based health education diabetes foot app to prevent serious foot complications: a feasibility study. *Practical diabetes* 35, 6 (2018), 203–209d.

[29] Oladapo Oyebode, Felwah Alqahtani, and Rita Orji. 2020. Using Machine Learning and Thematic Analysis Methods to Evaluate Mental Health Apps Based on User Reviews. *IEEE Access* 8 (2020), 111141–111158. doi:10.1109/ACCESS.2020.3002176

[30] Adam Palmquist. 2023. Design elements of conflict: A design study of a gamified smartphone application for employee onboarding. *Technology, Knowledge and Learning* 28, 3 (2023), 1133–1173.

[31] Zhongling Pi, Jianzhong Hong, and Jiumin Yang. 2017. Effects of the instructor's pointing gestures on learning performance in video lectures. *British Journal of Educational Technology* 48, 4 (2017), 1020–1029.

[32] Antoine Piau, Yoann Charlon, Eric Campo, Bruno Vellas, Fati Nourhashemi, et al. 2015. A smart insole to promote healthy aging for frail elderly individuals: specifications, design, and preliminary results. *JMIR rehabilitation and assistive technologies* 2, 1 (2015), e4084.

[33] Stefan Resch, André Kousha, Anna Carroll, Noah Severinghaus, Felix Rehberg, Marco Zatschker, Yunus Söyleyici, and Daniel Sanchez-Morillo. 2025. Smart Device Development for Gait Monitoring: Multimodal Feedback in an Interactive Foot Orthosis, Walking Aid, and Mobile Application. *Technologies* 13, 12 (2025). doi:10.3390/technologies13120588

[34] Stefan Resch, Lucian Zimmermann, Alexander Kunkel, Dennis Gerstung, Valentin Schwind, Diana Völz, and Daniel Sanchez-Morillo. 2024. Usability Evaluation of a Mobile Application for Foot Health Monitoring of Smart Insoles: A Mixed Methods Study. In *Adjunct Proceedings of the 26th International Conference on Mobile Human-Computer Interaction* (Melbourne, VIC, Australia) *(MobileHCI '24 Adjunct)*. Association for Computing Machinery, New York, NY, USA, Article 17, 7 pages. doi:10.1145/3640471.3680233

[35] Stefan Resch, Khalid Zoufal, Imad Akhouaji, Mohamed-Amin Abbou, Valentin Schwind, and Diana Völz. 2023. Augmented Smart Insoles – Prototyping a Mobile Application: Usage Preferences of Healthcare Professionals and People with Foot Deformities. *Current Directions in Biomedical Engineering* 9, 1 (2023), 698–701. doi:10.1515/cdbme-2023-1175

[36] Ljilja Ruzic, C Harrington, and J Sanford. 2017. Universal design mobile interface guidelines for mobile health and wellness apps for an aging population including people aging with disabilities. *Int J Adv Softw* 10, 3 (2017), 372–84.

[37] Clara Sayffaerth, Annika Köhler, Julian Rasch, Albrecht Schmidt, and Florian Müller. 2025. Through the Expert's Eyes: Exploring Asynchronous Expert Perspectives and Gaze Visualizations in XR. In *2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Daejeon, Korea, 1311–1321. doi:10.1109/ISMAR67309.2025.00136

[38] Martin Schrepp and Harry Budi Santoso. 2024. Construction of a Questionnaire to Measure the Learner Experience in Online Tutorials. In *Proceedings of the 20th International Conference on Web Information Systems and Technologies, WEBIST 2024 (International Conference on Web Information Systems and Technologies, WEBIST - Proceedings)*, Francisco Garcia Penalvo, Karl Aberer, and Massimo Marchiori (Eds.). Science and Technology Publications, Lda, Portugal, 315–322. doi:10.5220/0013014700003825 Publisher Copyright: Copyright © 2024 by SCITEPRESS.; 20th International Conference on Web Information Systems and Technologies, WEBIST 2024 ; Conference date: 17-11-2024 Through 19-11-2024.

[39] Valentin Schwind, Pascal Knierim, Cagri Tasci, Patrick Franczak, Nico Haas, and Niels Henze. 2017. "These are not my hands!": Effect of Gender on the Perception of Avatar Hands in Virtual Reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 1577–1582. doi:10.1145/3025453.3025602

[40] Alice Snell. 2006. Researching onboarding best practice: Using research to connect onboarding processes with employee satisfaction. *Strategic HR Review* 5, 6 (2006), 32–35.

[41] Brendan Strahm, Colin M. Gray, and Mihaela Vorvoreanu. 2018. Generating Mobile Application Onboarding Insights Through Minimalist Instruction. In *Proceedings of the 2018 Designing Interactive Systems Conference* (Hong Kong, China) *(DIS '18)*. Association for Computing Machinery, New York, NY, USA, 361–372. doi:10.1145/3196709.3196727

[42] Francesco Vona, Julia Schorlemmer, Michael Stern, Navid Ashrafi, Maurizio Vergari, Tanja Kojic, and Jan-Niklas Voigt-Antons. 2025. Comparing Pass-Through Quality of Mixed Reality Devices: A User Experience Study During Real-World Tasks. In *2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, Saint Malo, France, 1432–1433. doi:10.1109/VRW66409.2025.00362

[43] Pawel Weichbroth. 2025. Usability Issues With Mobile Applications: Insights From Practitioners and Future Research Directions. *IEEE Access* 13 (2025), 91301–91311. doi:10.1109/ACCESS.2025.3573503

[44] Laryssa Whittaker. 2025. Onboarding and offboarding in virtual reality: A user-centred framework for audience experience across genres and spaces. *Convergence* 31, 1 (2025), 87–106.

[45] Julia Woodward, Yu-Peng Chen, Katarina Jurczyk, Kathryn M. Ross, Lisa Anthony, and Jaime Ruiz. 2021. A Survey of Notification Designs in Commercial mHealth Apps. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 232, 7 pages. doi:10.1145/3411763.3451639

[46] Fei Wu, Jerald Thomas, Shreyas Chinnola, and Evan Suma Rosenberg. 2020. Exploring Communication Modalities to Support Collaborative Guidance in Virtual Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, Atlanta, GA, USA, 79–86. doi:10.1109/VRW50115.2020.00021

[47] Fu-Chia Yang, Pedro Acevedo, Siqi Guo, Minsoo Choi, and Christos Mousas. 2025. Embodied Conversational Agents in Extended Reality: A Systematic Review. *IEEE Access* 13 (2025), 79805–79824. doi:10.1109/ACCESS.2025.3566698

[48] Tingting Yang and Heng Luo. 2017. Learning Experience Questionnaire for Multimedia Case-Based Instruction: Development, Validation and Implications for Design. In *Blended Learning. New Challenges and Innovative Practices*. Springer, Springer International Publishing, Cham, 400–411.

[49] Chia-yin Yu. 2024. Enhancing Usability and User Experience Through User Onboarding in an Ecological Information App. In *Proceedings of the Future Technologies Conference*, Kohei Arai (Ed.). Springer, Springer Nature Switzerland, Cham, 376–389.

A Quantitative Results

A.1 System Usability Scale (SUS)

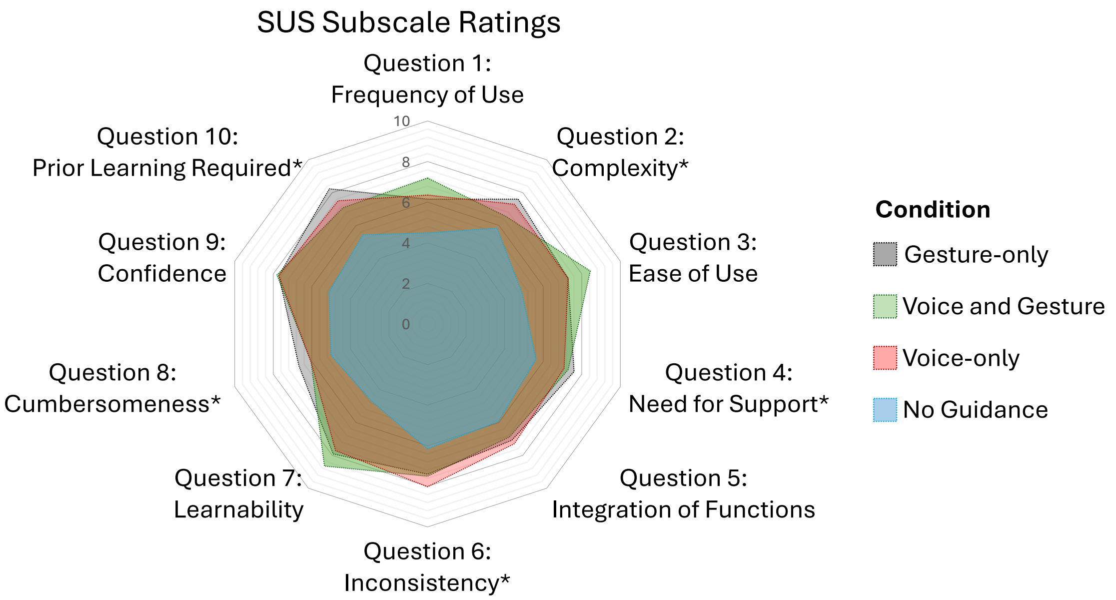


Figure 4: Radar plot of the ten SUS items comparing the four experimental conditions. Asterisk-marked items were reverse-scored to align scale directionality, such that higher values indicate more positive usability ratings.

Table 1: Significant effects (ART ANOVA) and post-hoc comparisons for SUS subscales.

SUS Item	Effect	Post-hoc comparisons
1: Frequency of Use	Gesture: $F(1, 69) = 4.97,\ p = .029$ Voice: $F(1, 69) = 5.51,\ p = .022$	non-visible < visible ($p = .056$) audible < non-audible ($p = .038$)
3: Ease of Use	Gesture: $F(1, 69) = 11.99,\ p < .001$ Voice: $F(1, 69) = 11.20,\ p = .001$	non-visible < visible ($p = .006$) audible < non-audible ($p = .008$)
4: Need for Support	Gesture: $F(1, 69) = 4.56,\ p = .036$	–
6: Inconsistency	Gesture×Voice: $F(1, 69) = 4.20,\ p = .044$	No Gesture: audible < non-audible ($p = .012$)
7: Learnability	Gesture: $F(1, 69) = 11.77,\ p = .001$ Voice: $F(1, 69) = 11.11,\ p = .001$ Gesture×Voice: $F(1, 69) = 4.23,\ p = .043$	No Voice: non-visible < visible ($p < .001$) No Gesture: audible > non-audible ($p < .001$)
9: Confidence	Gesture: $F(1, 69) = 7.35,\ p = .008$ Voice: $F(1, 69) = 7.97,\ p = .006$ Gesture×Voice: $F(1, 69) = 7.32,\ p = .009$	No Voice: non-visible < visible ($p = .002$) No Gesture: audible > non-audible ($p = .002$)
10: Prior Learning	Gesture: $F(1, 69) = 5.01,\ p = .028$ Gesture×Voice: $F(1, 69) = 7.55,\ p = .008$	No Voice: non-visible > visible ($p = .003$) No Gesture: audible < non-audible ($p = .028$)

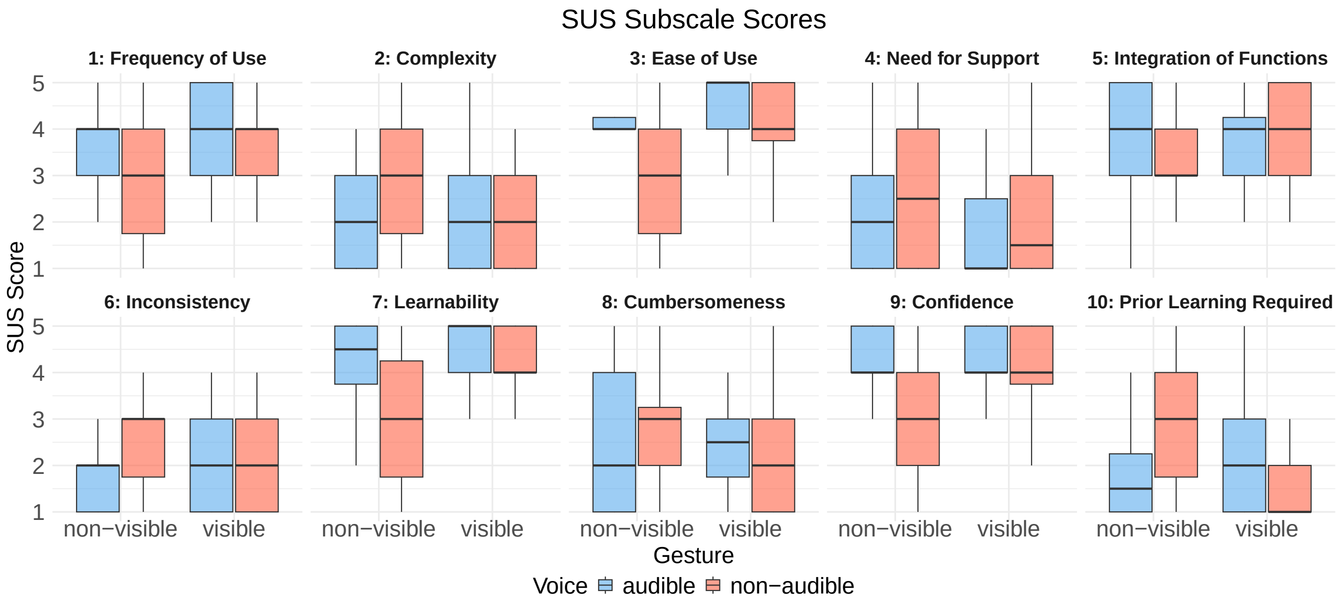


Figure 5: Boxplots of the ten SUS items. Ratings range from 1 (strongly disagree) to 5 (strongly agree).

A.2 Learning Experience Questionnaire (LEQ)

Table 2: Overview of significant main and interaction effects for LEQ subscales with Bonferroni-corrected post-hoc t-tests.

Subscale	Effect	F(1,23)	p	η_p^2	Post-hoc comparisons
Contextuality	Gesture	10.15	.004	.31	–
	Voice	13.28	.001	.37	–
	Gesture×Voice	5.41	.029	.19	No Voice: non-visible < visible ($p < .001$)
					No Gesture: audible > non-audible ($p < .001$)
Effects	Gesture	5.42	.029	.19	–
	Voice	7.72	.011	.25	–
	Gesture×Voice	11.78	.002	.34	No Voice: non-visible < visible ($p < .001$)
					No Gesture: audible > non-audible ($p < .001$)
Design Features	Voice	20.14	< .001	.47	–
	Gesture×Voice	6.49	.018	.22	No Voice: non-visible < visible ($p = .011$)
					No Gesture: audible > non-audible ($p < .001$)

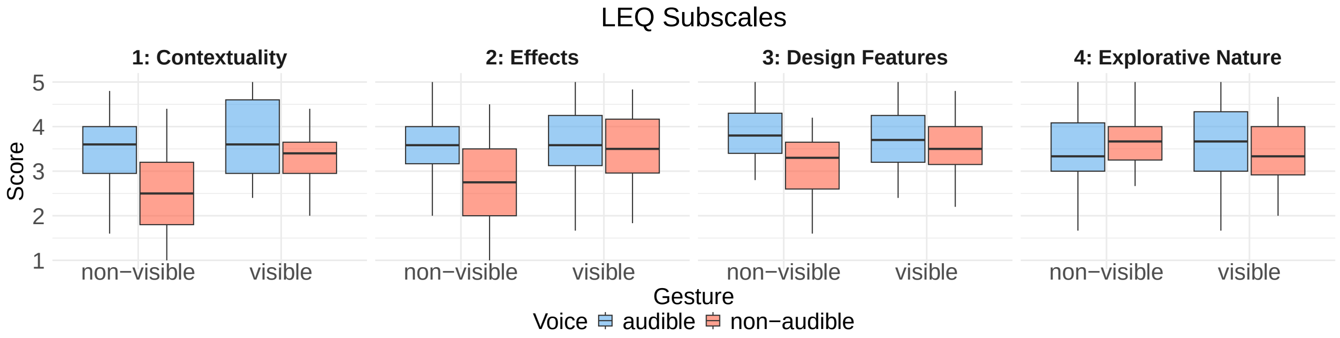
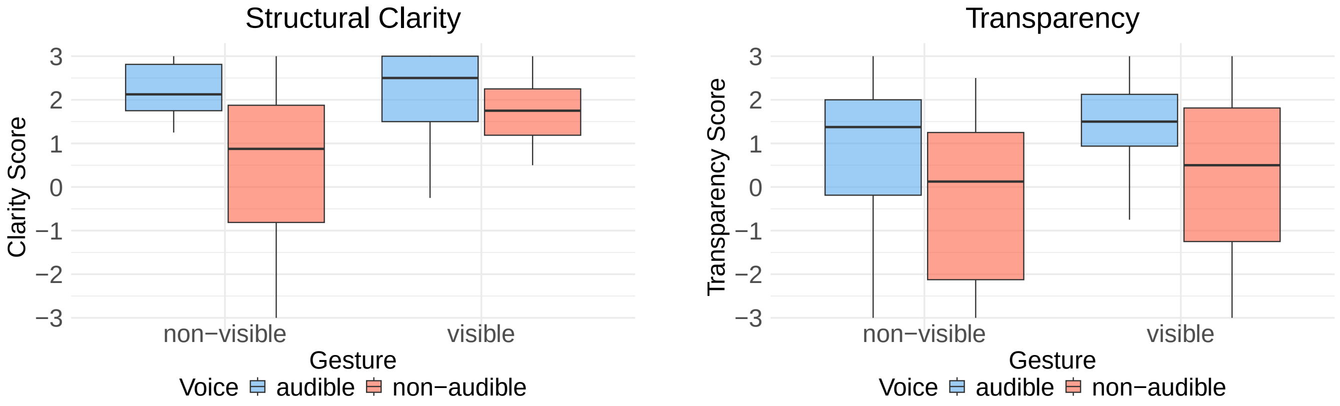


Figure 6: Boxplots of the four LEQ subscales. Ratings range from 1 (fully disagree) to 5 (fully agree).

A.3 Tutorial Experience Questionnaire (TEQ)



(a) Subscale "Structural Clarity"

(b) Subscale "Transparency"

Figure 7: Boxplots of the two TEQ subscales. Ratings range from -3 (strongly disagree) to +3 (strongly agree), with 0 as a neutral response.