Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 1119–1128
https://doi.org/10.54941/ahfe1008342

The Role of Mental Models in Learning How to Use Smart Glasses With Display Among First Time Users

Alan Turatti, Marcos Silbermann, Randerson Queiroz, Rodrigo Cavalcanti, and Hanah Correa

Sidia Institute of Science and Technology, Av. Darcy Vargas, 654, 69055-035 Manaus, Brazil

ABSTRACT

Smart glasses with display introduce interaction methods that remain unfamiliar to most users, requiring them to interpret the experience through piror technological knowledge. This study investigates how first time users construct, adapt, and consolidate their understanding while learning to use this emerging category of wearable interface. Ten participants took part in two sequential sessions involving free exploration, guided tutorials, and practical tasks using the device's touchpad and wristband. Findings show that participants relied on established mental models from smartphones, smartwatches, voice assistants, mice, joysticks, and remote controls to interpret the device and decide how to interact. However, understanding an interaction conceptually did not immediately translate into consistent practical execution. Over time, participants also shifted from initially favoring the more familiar touch interface to preferring the wristband for more complex tasks. The findings suggest that learnability in smart glasses depends on how interaction models relate to prior knowledge and on how effectively the system supports users in building new interaction repertoires.

Keywords: Mental models, Learnability, Exploratory learning, Onboarding, Knowledge transfer, Smart glasses, Technology adoption

INTRODUCTION

Smart glasses with display represent an emerging category of wearable interfaces that introduces interaction paradigms still unfamiliar to most users. Unlike devices such as smartphones and smartwatches, whose basic interaction patterns have become widely established, smart glasses may combine multiple input methods and small visual interfaces that create distinct interaction challenges (Ro et al., 2018). Learning to use these devices therefore involves not only discovering available commands, but also understanding how different interactions relate to the system and when each should be used. Comparatively little is known about what happens after adoption—how first-time users actually learn to operate these devices (Koutromanos and Kazakou, 2023).

When encountering an unfamiliar technology, users do not approach it without prior references. Instead, they form and apply mental models based on previous knowledge and experience to predict how a system works and

Received November 7, 2026; Revised November 19, 2026; Accepted November 22, 2026; Available online December 1, 2026

© 2026 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

determine how to interact with it (Norman, 1983; Johnson-Laird, 1983). In human-computer interaction, these models are not necessarily complete or technically accurate, but they allow users to interpret system behaviour and guide their actions (Norman, 1983). Previous knowledge can therefore facilitate the learning of a new interface when familiar interaction patterns can be transferred to it, while differences between expected and actual system behaviour may create confusion or require existing models to be adapted (Singley and Anderson, 1989).

This relationship is particularly relevant for smart glasses with display. Familiar elements, such as physical buttons and touch interactions, coexist with less established methods such as gesture-based input through a wrist-worn device. Users may therefore draw on interaction patterns learned from other technologies to make sense of the device, while simultaneously developing new understanding for interactions that have no clear equivalent in their previous experience. From a learnability perspective, this process extends beyond immediate task completion: learning also involves developing the ability to use a system effectively over time and after continued or repeated experience (Grossman et al., 2009).

This article presents the results of a test with first-time users. The experiment was structured to highlight the learning processes of basic and advanced device interactions, considering the different input methods available. Their results fuel our discussion about how first time users construct, adapt, and consolidate their understanding while learning to use smart glasses with display. By examining the relationship between prior mental models and the development of new interaction repertoires, the study aims to contribute to the understanding of learnability and onboarding in emerging wearable interfaces.

MENTAL MODELS AND PRIOR KNOWLEDGE

Mental models are internal representations that people construct to understand and interpret technologies and interactive systems, guiding how they interact with them (Johnson-Laird, 1983; Norman, 2013). These representations do not necessarily correspond to the expected functioning of the technology. Instead, they constitute interpretations, often incomplete and subject to inconsistencies, about the functioning of the technological device, being continuously refined as new experiences are incorporated (Norman, 2013). Because they influence how users interpret unfamiliar technologies and build strategies to learn how to use them, mental models play a central role in learning new interfaces.

When interacting with a technology for the first time, people rarely start from scratch. They draw on knowledge, experiences, and expectations built up from previous interactions to interpret how a new system works and how it should be used. This process allows them to formulate initial hypotheses about possible actions and outcomes, even in the absence of familiarity with the technology in question (Norman, 2013). When these expectations match the system's behavior, learning tends to occur more smoothly. On the other hand, when there are discrepancies between the user's mental model and the

conceptual model adopted by the system, misinterpretations, difficulties in use, and the need for adaptation may arise (Norman, 2013).

The updating of mental models occurs progressively as new experiences provide evidence about how the system works. During this process, users confirm, adjust, or abandon initial interpretations, gradually refining their understanding of the technology (Johnson-Laird, 1983; Norman, 2013). In emerging technologies, where there are few established usage references, this process tends to depend even more on the transfer of knowledge from already familiar devices and interfaces. Understanding how these mental models are constructed, used, and modified is therefore fundamental to understanding how new users learn to use unfamiliar interfaces, such as smart glasses with displays.

In the case of Meta Ray-Ban Display, this transfer of repertoire is intense. The interaction channels are distributed among physical buttons on the frame, a touchpad on the temple, and a wristband that recognizes gestures based on wrist muscle activity (Kaifosh et al., 2025), while the response to these actions appears on a display positioned in the field of vision. Each of these elements refers to distinct references: the touchpad refers to smartphones and notebooks; the buttons, to electronic devices in general; and the wristband has no direct analogies. It is this dynamic between forms of input with little consistency with the users' repertoire and a visual interface that does not respond on the same plane as the action is performed that the present study investigates, asking what prior references of mental models first-time users mobilize to recognize each input and how these references relate to what they see on the display.

LEARNING AND RETENTION THROUGH INTERACTION

Learning a new interface involves more than just discovering commands or functionalities. As users interact with a system, they build an understanding of its operating logic, refine their mental models, and develop greater confidence in performing tasks. This process occurs gradually, being influenced by practice, feedback provided by the interface, and opportunities to apply the acquired knowledge in different contexts (Nielsen, 1993).

In emerging technologies, such as smart glasses with displays, this process can be even more challenging. Unlike widely used devices such as smartphones and computers, smart glasses introduce interaction methods and ways of use that are still unfamiliar to most users. In addition, they often combine multiple input modalities, such as touch, voice, gestures, and wearable sensors, requiring the user to understand not only how to perform each interaction, but also when each method is most appropriate for the situation. The absence of widely established usage conventions leads users to rely on expectations built from already known technologies to interpret how these devices work; this phenomenon is known as legacy bias (Morris et al., 2014; Wobbrock et al., 2009).

Beyond the experience of initial contacts, ease of learning also encompasses what users retain after a period without contact with the system. Nielsen (1993) calls this memorability. The way interactions are initially learned also influences their retention. Preparing users with familiar cultural references

may improve the ease of learning and memorability of gestural interactions (Ali, Morris, and Wobbrock, 2021). Furthermore, interfaces that require mental effort during learning, making users actively retrieve information instead of just following visual instructions, generate more consolidated spatial and gestural learning (Cockburn et al., 2007). Taken together, these findings suggest that tutorials that demonstrate interactions without stimulating active practice may favor understanding but result in weak retention.

In smart glasses with displays, this perspective suggests that retention is not evenly distributed across the layers of learning. Recognition of the device and its input methods can be established from the first contact, while the vocabulary of commands remains dependent on reinforcement, especially ambiguous gestures. Discerning these layers is what allows us to separate what onboarding needs to teach only once from what needs to be sustained throughout use. Our study sought to map these learning processes in order to identify areas for improvement in usability and potential resistance these devices may encounter regarding user retention.

METHODS AND TOOLS

Ten people participated in the study, selected for having little or no prior experience with smart glasses and other extended reality (XR) devices. This criterion aimed to reduce the influence of mental models formed by the previous use of technologies with characteristics and forms of interaction like those of the smart glasses investigated.

The study was structured in two sessions, held approximately one week apart, to observe learning from the first contact to the resumption of use. It is important to emphasize that the tests were structured for qualitative observation of use without measuring the usability of the device. The device used was the Rayban Meta Display, considered state-of-the-art in this type of technology. This product offers different interaction methods, including physical buttons, touchpad, and gestures performed via a neural wristband. Each phase had a specific test script, which included direct experimentation and in-depth interviews; the sessions lasted about an hour.

Phase 1: The script for this phase was structured with an emphasis on the learning experience of basic interactions through the use of the touchpad and neural band. The steps included the out-of-the-box moment, initial free experimentation, use of the tutorial and minigames available through the system and aimed at learning basic interactions. In the final stage of this script, participants were asked about difficulties and their opinion about the experience.

Phase 2: This phase of the experiment focused on the connection between memorizing the basic interactions learned in the previous session and learning complex interactions. The script began with questions about the participants' memories of the interactions from the previous phase, followed by a free experimentation phase in which most participants sought to repeat the interactions performed. In subsequent phases, participants performed the complex interactions available in the device's tutorial, followed by in-depth interviews.

DATA COLLECTION AND ANALYSIS

During the tests, we made audio and video recordings of the sessions, in addition to notes taken by the researchers involved in conducting the test—at least two researchers per session. Subsequently, the data were tabulated and checked with the support of the video recordings for later analysis of the interview content and quantification of the difficulties encountered by the users, with contextual support from the researchers' notes. The results are presented in the next section of this article.

RESULTS: BUILDING AND ADAPTING UNDERSTANDING FOR SMART GLASSES

The initial contact with smart glasses was marked by a gradual process of building understanding about the device and its interaction methods. The exploration focused primarily on the visible components of the device. Physical buttons, touchsensitive surfaces, cameras, and the wristband itself served as the first clues used to formulate hypotheses about its functioning. At this point, understanding was still heavily influenced by previous experiences with other technologies. Instead of approaching smart glasses as a completely new device, participants attempted to interpret them based on already established mental models, drawing analogies with smartphones, smartwatches, voice assistants, mousepads, remote controls, etc. These references guided not only how participants understood the device but also their expectations about how they should interact with it. In some cases, these expectations led to behaviors such as trying to talk directly to the wristband, making gestures in space, or positioning the wristband on the non-dominant arm, reproducing usage patterns associated with other technologies.

As exploration progressed, these hypotheses were constantly confronted with the system's responses. The tutorial reduced initial uncertainty by explaining input methods that free exploration had not revealed, such as bracelet gestures, which have no physical indication of their existence. However, participants completed the demonstration ready to describe the interaction and yet still failed to execute it, especially when execution depended on motor precision or spatial mapping, rather than comprehension.

At the end of the first session, a significant change was noticeable in how participants interpreted the device. Understanding ceased to depend exclusively on analogies with previously known technologies and began to incorporate experiences built during the interaction with the smart glasses themselves. Although this mental model was still under development and did not represent complete mastery over all interactions, it provided a more consistent basis for understanding the overall functioning of the system and guiding the actions performed throughout the experience.

The physical buttons represented the first point of interaction explored by the participants. Even before exploring the touchpad or the wristband, many participants sought to identify how the glasses were turned on. However, there was frequent confusion between the capture button and the side switch, responsible for turning the device on and off. Faced with this

uncertainty, participants resorted to feedback provided by the system, such as the Meta logo and startup beeps, to confirm that the glasses had been turned on correctly. In one case, the attempt resulted in the accidental reset of the device, highlighting that, although physical elements evoked familiar mental models, their function was not always clearly communicated.

The touchpad was recognized by participants as a familiar interaction surface, often associated with smartphones and notebook touchpads. However, this familiarity did not always facilitate its use. Several participants tried to apply already known navigation patterns, but found it difficult to adapt these movements to the physical orientation of the handle, where sliding forward or backward corresponded to navigating between options presented laterally on the interface. Two participants reported that the extent of the touch-sensitive surface was not legible: one described the temple as if it were divided into sectors, and another judged the area too small for the gestures it was expected to receive. Although the device had a relief delimiting this surface, its function was not readily understood, contributing to interaction errors during the initial learning.

On the other hand, the neural band presented the steepest learning curve among the three input methods. In the first contact, it wasn't even recognized as a means of interaction: without any signifier indicating its function, it remained invisible until presented in the tutorial. By the second session, eight of the ten participants stated a preference for the wristband, one for the temple touchpad and one reported no clear preference. Observed use, however, did not follow a single pattern: the wristband was the channel participants reached for first and the only one used for zooming and image manipulation, while the touchpad remained a physical anchor and a fallback. The most remarkable aspect of this change is that it didn't depend on complete memorization: participants who didn't remember the exact gesture of a specific command still chose the band, relying on the certainty that the gesture existed and could be retrieved.

Although most interactions were assimilated throughout the two sessions, some commands continued to cause recurring confusion. The back and sleep mode activation commands, for example, were frequently confused by participants. Confusion was also observed between other commands, such as volume and zoom adjustment gestures, indicating that small differences between gestures continued to create difficulties during learning.

In the second phase of the experiment, participants went through stages of free use of the device and answered questions about remembering the interactions performed in the first phase, in addition to stages related to learning complex interactions. Understanding the general functioning of the device proved to be more stable than remembering specific commands and gestures.

Retention did not fail evenly across commands. Volume, executed on the wristband by pinching thumb and index finger and rotating the wrist, was reproduced by eight of ten participants after the one-week interval. The same motor form, however, is required to zoom: in six of ten cases the volume control was activated while the declared intention was to enlarge an image, and two further participants failed at an earlier step of the flow.

Only two completed the zoom. The form of the gesture therefore survived the interval more reliably than the function it was meant to serve—the pattern Nacenta et al. (2013) describe for gestural vocabularies, and one that here follows from two functions sharing a single motor form rather than from any difficulty in execution.

Recovery was also observed. Three participants returned to the system tutorial during the second session to relearn the zoom gesture and executed it successfully on the following attempt, indicating that what had decayed was retrieval of the association rather than the capacity to perform the gesture.

DISCUSSION

This study demonstrated how users learn during their first experience with smart glasses, focusing on how they understand and retain both basic and complex interactions. It also showed that, throughout the learning process of this specific technology, users construct an interaction vocabulary by updating a repertoire inherited from technologies they have previously used as they gradually discover how to operate this new device. Describing smart glasses learning in this way makes it possible to explain, through a single mechanism, three phenomena that appeared separately in the results: where users' initial interaction hypotheses come from, why understanding an interaction is not sufficient to perform it correctly, and what remains available after a period without use. This perspective also helps anticipate usage patterns as well as potential usability problems and frustrations that the initial experience may generate, thereby limiting user engagement.

Participants interpreted the interactions available on the device by drawing on references from smartphones, smartwatches, voice assistants, computer mice, joysticks, and remote controls, and their first interaction attempts were concentrated on the physical buttons. This finding is expected because the button is the only element of the device whose possible action is physically signaled—it constitutes an affordance with a signifier, according to Norman's (2013) distinction. The only signifier that effectively operated in this study was auditory rather than visual: participants located both the buttons and the touch-sensitive surface through the sounds emitted when they were pressed or touched. Wherever the device provided a signal, exploration was successful; where it did not, the interaction channel remained outside the user's vocabulary until it was explicitly taught. The remaining interaction channels provide no perceptible cues and therefore do not become part of the interaction vocabulary through exploration alone, but only through instruction. What elicitation studies describe as *legacy bias* (Morris et al., 2014; Wobbrock, Morris, & Wilson, 2009) appears here not as a response bias in an experimental task, but as the very material from which the initial interaction vocabulary is necessarily constructed. The raised contour delimiting the touch-sensitive area on the temple provides a counterexample within the same distinction: it exists as an affordance without functioning as an effective signifier (Norman, 2013). Although physically present, it failed to communicate its function, leading participants to perform gestures outside the interactive area.

In smart glasses, the combination of the temple touchpad, the neural band, and the display forms a particularly unstable interaction system because of the difficulty of providing cues that introduce and reinforce elements of users' inherited interaction vocabulary. In this sense, the temple touchpad was recognized most rapidly because it has an obvious counterpart in users' previous experience: the touch-sensitive surfaces of smartphones and laptops. However, it also generated a variety of functional difficulties by inducing false correspondences with touch surfaces found on other devices. Prior experience with smartphones and laptops associates gestures with the plane on which they are performed, whereas on the temple the gesture is executed along the horizontal axis while its outcome appears along the display's vertical axis. These competing mappings require users to suppress the previously learned one before the correct mapping can be acquired. As argued by Schumacher and Gentner (1988), procedural transfer may hinder rather than facilitate performance. Consequently, the temple touchpad enters the interaction vocabulary early but remains unstable within it.

The neural band exhibits the opposite pattern. It has no corresponding interaction in users' prior repertoire and provides no perceptible indication of its function, keeping it outside the interaction vocabulary until it is explicitly taught. Once incorporated, however, it is learned without interference from previously acquired mappings because there is no existing mapping to suppress or replace. This is the most plausible explanation for the migration observed across sessions, in which participants progressively adopted the neural band for more complex tasks.

Confidence in the interaction method, therefore, was consolidated before mastery of the interaction vocabulary. This pattern differs from that observed with the temple touchpad, where initial familiarity did not translate into equivalent confidence. Retention failures were primarily associated with the mapping between commands and gestures rather than with gesture execution itself. Participants correctly reproduced the gestures but applied them to the wrong functions or contexts, a pattern consistent with Nacenta et al.'s (2013) description of gestural vocabularies, in which forgetting results primarily from errors in the association between commands and gestures.

CONCLUSION

This study examined how first-time users construct, adapt and consolidate their understanding while learning to use smart glasses with display. Observed across two sessions one week apart, the process is better described as the construction of an interaction vocabulary than as the discovery of a fixed set of commands: participants assembled that vocabulary from the repertoire inherited from other technologies and revised it through each attempt they executed.

The study is exploratory and qualitative, conducted with ten participants and a single device, and the two sessions do not allow the point at which the interaction vocabulary stabilises to be observed. Future work should follow the construction of this vocabulary over longer periods and measure, through a savings paradigm, the residual retention that resumption after an interval only suggests.

ACKNOWLEDGMENT

This paper was presented as part of the results of the project "SIDIAM_VST_PLATFORM_AND_APPLICATIONS", carried out by the Institute of Science and Technology – SIDIA, in partnership with Samsung Eletrônica da Amazônia LTDA, in accordance with the Information Technology Law n.8387/91 and article ate the.39 of Decree 10,521/2020.

REFERENCES

Ali, A.X., Morris, M.R., Wobbrock, J.O.: "I am Iron Man": Priming improves the learnability and memorability of user-elicited gestures. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Article 359, pp. 1–14. ACM, New York (2021). doi: 10.1145/3411764.3445758.

Cockburn, A., Kristensson, P.O., Alexander, J., Zhai, S.: Hard lessons: Effort-inducing interfaces benefit spatial learning. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1571–1580. ACM, New York (2007). doi: 10.1145/1240624.1240863.

Grossman, T., Fitzmaurice, G., Attar, R.: A survey of software learnability: Metrics, methodologies and guidelines. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 649–658. ACM, New York (2009). doi: 10.1145/1518701.1518803.

Johnson-Laird, P.N.: Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness. Harvard University Press, Cambridge, MA (1983).

Kaifosh, P., Reardon, T.R., CTRL-labs at Reality Labs: A generic non-invasive neuromotor interface for human-computer interaction. Nature 645(8081), 702–711 (2025). doi: 10.1038/s41586-025-09255-w.

Koutromanos, G., Kazakou, G.: Augmented reality smart glasses use and acceptance: A literature review. Computers & Education: X Reality 2, 100028 (2023). doi: 10.1016/j.cexr.2023.100028.

Morris, M.R., Danielescu, A., Drucker, S., Fisher, D., Lee, B., Schraefel, M.C., Wobbrock, J.O.: Reducing legacy bias in gesture elicitation studies. Interactions 21(3), 40–45 (2014). doi: 10.1145/2591689.

Nacenta, M.A., Kamber, Y., Qiang, Y., Kristensson, P.O.: Memorability of pre-designed and user-defined gesture sets. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1099–1108. ACM, New York (2013). doi: 10.1145/2470654.2466142

Nielsen, J.: Usability Engineering. Academic Press, San Diego (1993).

Norman, D.A.: Some observations on mental models. In: Gentner, D., Stevens, A.L. (eds.) Mental Models, pp. 7–14. Lawrence Erlbaum Associates, Hillsdale, NJ (1983).

Norman, D.A.: The Design of Everyday Things: Revised and Expanded Edition. Basic Books, New York (2013).

Ro, Y.K., Brem, A., Rauschnabel, P.A.: Augmented reality smart glasses: Definition, concepts and impact on firm value creation. In: Jung, T., tom Dieck, M.C. (eds.) Augmented Reality and Virtual Reality: Empowering Human, Place and Business, pp. 169–181. Springer, Cham (2018). doi: 10.1007/978-3-319-64027-3_12.

Schumacher, R.M., Gentner, D.: Transfer of training as analogical mapping. IEEE Transactions on Systems, Man, and Cybernetics 18(4), 592–600 (1988). doi: 10.1109/21.17377.

Singley, M.K., Anderson, J.R.: The Transfer of Cognitive Skill. Harvard University Press, Cambridge, MA (1989).
Wobbrock, J.O., Morris, M.R., Wilson, A.D.: User-defined gestures for surface computing. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1083–1092. ACM, New York (2009). doi: 10.1145/1518701.1518866.