

The Pen: Episodic Cognitive Assistance via an Ear-Worn Interface

YONATAN TUSSA, University of Maryland, College Park, United States

ANDY HEREDIA, University of Maryland, Global Campus, United States

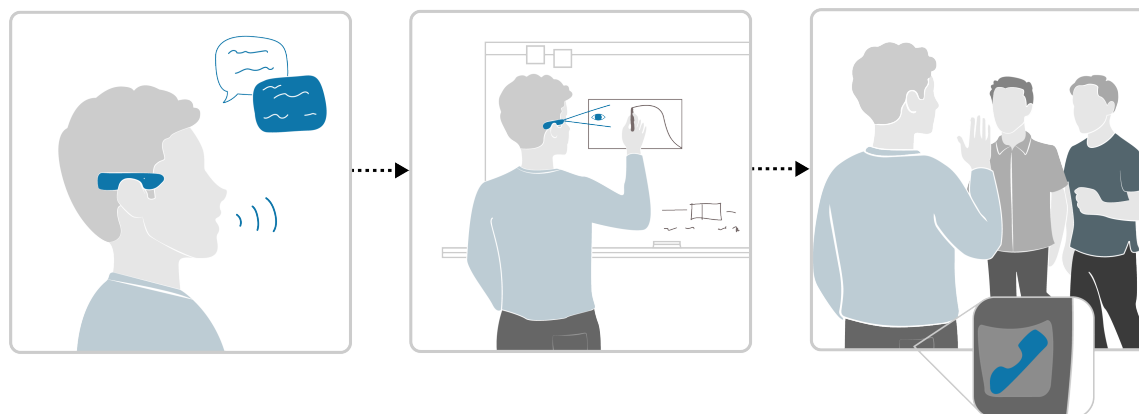


Fig. 1. The Pen is worn behind the ear and the user is speaking to it (left); user continues whiteboarding while the device observes visual context (middle); user disengages by removing the device and placing it in a pocket before resuming social interaction (right).

Wearable AI is often designed as always-available, yet continuous availability can conflict with how people work and socialize, creating discomfort around privacy, disruption, and unclear system boundaries. This paper explores episodic use of wearable AI, where assistance is intentionally invoked for short periods of focused activity and set aside when no longer needed, with a form factor that reflects this paradigm of wearing and taking off a device between sessions. We present The Pen, a ear-worn device resembling a pen, for situated cognitive assistance during tasks where a user's hands may be busy. The device supports short, on-demand assistance sessions using voice and visual context, with clear start/end boundaries and local processing without sending personal data to the cloud. We report findings from an exploratory study examining how users interpret episodic activation, agency, and social legibility in wearable AI.

CCS Concepts: • **Human-centered computing**; • **Information systems** → **Mobile information processing systems**; • **Hardware** → **Communication hardware, interfaces and storage**; **Emerging technologies**;

Additional Key Words and Phrases: Wearable computing, Earables, Voice User Interfaces, Interaction design, Legibility, Intelligibility

ACM Reference Format:

Yonatan Tussa and Andy Heredia. 2026. The Pen: Episodic Cognitive Assistance via an Ear-Worn Interface. 1, 1 (January 2026), 6 pages. <https://doi.org/XXXXXXX.XXXXXXX>

Authors' Contact Information: Yonatan Tussa, University of Maryland, College Park, College Park, United States, ytussa@umd.edu; Andy Heredia, aheredia7@umgc.edu, University of Maryland, Global Campus, College Park, United States.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2026 Copyright held by the owner/author(s). Publication rights licensed to ACM.

Manuscript submitted to ACM

Manuscript submitted to ACM

1 Introduction

Wearable computing promises in-the-moment assistance during everyday tasks, yet most wearable AI systems assume continuous availability—always listening, sensing, and ready for interaction. In practice, persistent sensing raises privacy concerns for wearers and bystanders, introduces ambiguity around when data is being collected, and can undermine users’ sense of agency and control over the system’s operation. [1, 3–6, 11, 18].

A recurring limitation of prior wearable AI systems is their assumption that intelligence should operate continuously in the background. These designs clash with important boundaries: when the system is active, what it may access, and how its involvement is legible to others. As a result, systems can feel intrusive. Existing interaction paradigms offer limited relief, with smartphones and laptops requiring shifts in gaze and manual attention that can disrupt cognitive flow. Voice assistants reduce visual demand but can be socially exposed, especially in shared workspaces. At the same time, ear-worn platforms are becoming increasingly capable sensing and interaction devices, enabling inference from audio and motion signals in everyday form factors [10, 20].

Rather than defaulting to ambient, always-on intelligence, we explore episodic wearable assistance where users wear a device for short, task-bounded sessions and disengage when the task ends. This approach takes inspiration from foundational HCI work on intelligibility and accountability in context-aware systems [4], and on research showing that the act of wearing itself shapes expectations, meaning, and social interpretation [14, 18]. This aligns with a growing body of work on proactive and assistive agents that must carefully manage when and how they intervene [16, 17].

We investigate this paradigm with The Pen, an ear-worn wearable designed for short, task-bounded assistance sessions. Users put the device behind their ear at the start of a task and remove it when the task ends, making activation boundaries legible through physical interaction.

Our contribution centers on interaction design, focusing on how form factor, activation rituals, and feedback mechanisms shape what cognitive assistance feels like in real-world settings.

We ask the following questions:

- How do users initiate and end assistance without breaking flow?
- What cues help users understand what the system is doing in the moment?
- How do episodic boundaries shape users’ sense of agency, control, and cognitive flow?

Our contributions are the following:

- **Exploration:** We explore episodic use of a wearable AI assistant, where intentional wearing marks the beginning and end of short assistance sessions.
- **System:** We present The Pen, a pen-shaped ear-worn interface designed to support in-the-moment assistance.
- **Study:** We report insights into episodic wearable interaction and tensions between agency, legibility, and social comfort.

Relationship to Prior Work. This paper builds on our prior work [24], which examined the systems, hardware integration, and privacy-preserving challenges involved in building a multimodal wearable capable of local voice-and-vision inference. Here we focus on interaction design and user experience of episodic, task-bounded assistance.

2 Related Work

Our work draws on research in ear-worn and screenless wearables, situated cognitive assistance, and the social and ethical implications of wearable sensing. Across these areas, many systems assume persistent availability and continuous inference. In contrast, we focus on episodic use: assistance invoked for short, task-bounded moments and then disengaged, with interaction boundaries made legible through form factor and feedback.

2.1 Ear-Worn and Screenless Wearable Interaction

Prior work explores discreet, hands-free interaction through head and ear-worn devices. AlterEgo demonstrates silent-speech interaction as a pathway toward always-available computing [12]. Surveys and taxonomies of earables highlight the ear as a socially acceptable site for audio-first interaction while surfacing challenges around comfort, ergonomics, sensing reliability, and legibility [8, 20]. Recent work shows that commodity earphones can support inference beyond audio I/O (e.g., head motion tracking via fused IMU and acoustics) [10]. EarRumble further show how subtle bodily actions can enable hands and eyes-free control [19].

Beyond earables, screenless wearables such as FingerTrak and IRIS demonstrate lightweight capture of task-relevant context without the user shifting attention to a phone [9, 13]. FingerTrak shows how a wearable can infer task-relevant hand activity under partial visibility using sensing to recover interaction context and IRIS demonstrates a camera-enabled ring that captures brief visual context and offloads recognition to a paired device.

2.2 Situated Assistance and Memory Support

A growing body of work explores AI assistance embedded within ongoing activities. Systems such as Memoro use large language models to provide concise, real-time memory augmentation during work [25], while other agents intervene during procedures or everyday tasks to offer guidance or corrective feedback [2, 16]. Related work on proactive conversational agents examines when and how systems should initiate support to remain helpful without becoming intrusive [17].

2.3 Activation Tax and Micro-Interaction Costs

Even when assistance is useful, the cost of invoking it can determine whether it fits into real work. Prior work quantifies the “activation tax” of mobile interaction, showing that simply initiating phone use can impose meaningful time and attention overhead, making micro-interactions sensitive to seconds-level friction [22]. This motivates our emphasis on episodic, wearable entry/exit: if users seek short bursts of help (e.g., clarification during reading or a quick capture during an incident), the interface must enable rapid engagement and disengagement without requiring sustained posture shifts or repeated retries.

2.4 Privacy, Social Acceptability, and Legible Boundaries

Wearable sensing raises longstanding concerns around bystanders, ambiguity, and the social implications of capture and inference [1, 6, 18]. Studies of camera-equipped wearables show that unclear system state and ambiguous recording boundaries can undermine trust and social comfort for both wearers and bystanders [5, 14]. Cardea further argues that camera privacy is context-dependent and emphasizes mechanisms that allow users to manage capture and sharing in ways that match situational expectations [21]. Technical approaches such as ScreenAvoider demonstrate methods for limiting unintended capture in ubiquitous sensing systems [15].

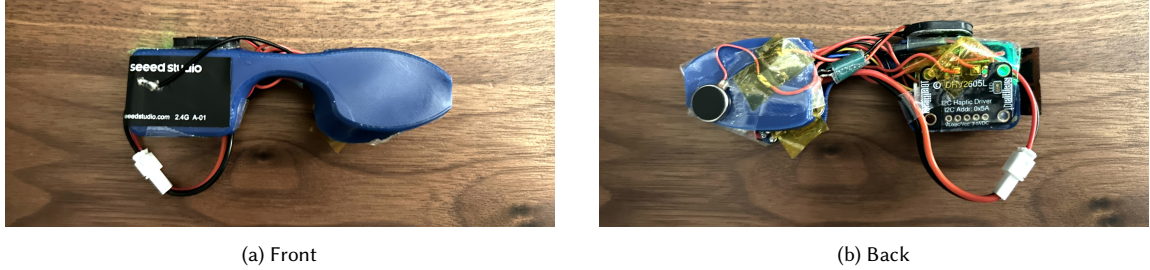


Fig. 2. Front and back of The Pen prototype.

Mirai describes an always-available wearable that continuously senses and delivers proactive assistance, illustrating both the promise of persistent inference and the risks of diminished legibility when systems are continuously listening and observing [7].

3 The Pen

The Pen is a wearable device that resembles a pen or pencil worn behind the user's ear, designed for hands-free interaction during focused work episodes. The device consists of a microphone, camera, speaker, Force Sensitive Resistor (FSR), haptic motor, and UFL antenna. The haptic motor and force sensor enable sensing and feedback for pressing the device to take a picture and a long-press gesture to initiate queries. Wearing and removing the device marks the beginning and end of an assistance episode. During a session, users interact through voice while remaining engaged in the task (e.g., brainstorming at a whiteboard, washing dishes, automobile maintenance, furniture assembly), with the device capturing visual context and generating task-relevant support such as clarifications or summaries. Additional technical details are described in prior work [24].

4 Study

We conducted an exploratory study with 6 participants (4 male, 2 female; ages 18-53) across two tasks: (1) reading/studying and (2) whiteboarding/brainstorming. Sessions lasted 12–30 minutes. Participants were instructed to treat wearing the device as the start of an assistance episode and removal as its end. Immediately after each session, we collected surveys and open-ended reflections to understand perceptions of agency, flow, and social comfort.

5 Discussion

Our findings show how episodic assistance in a wearable form factor shapes perceptions of agency, trust, flow, and social comfort. Across participants, The Pen was experienced as promising but not without friction, revealing tensions between intentional activation and desires for seamless intelligence.

5.1 Episodic Boundaries and Agency

Participants did not consistently interpret wearing the device as the beginning of an assistance episode. One participant noted that the episode only felt active after pressing the device, suggesting that physical wearing alone was insufficient to signal activation. Participants implicitly sought additional cues such as sounds or haptic feedback to clarify when an episode had begun. These observations indicate that episodic interaction requires explicit feedback to stabilize

user understanding of system state. In contrast, removing the device was consistently perceived as a clear end to the interaction: taking it off felt definitive and relieved any sense of ongoing system involvement.

Participants also experienced episodic use as increasing their sense of agency and intentional control over the system. Several preferred using the device only at selected moments rather than continuously, describing it as something that felt “extra” rather than always necessary. Voice interaction was perceived as natural and expressive, enabling participants to articulate questions without shifting attention to a phone and reducing reliance on manual interaction. Non-native English speakers perceived particular value in voice interaction, noting that handheld interfaces often introduce friction when interacting with systems designed primarily in a non-native language.

However, this sense of control came with interaction overhead. Participants reported friction when invoking the device, especially when visual capture required repeated attempts. Technical limitations occasionally disrupted task flow, with sensing failures forcing participants to pause their activity and reducing their ability to remain focused in some cases. As a result, the device was perceived as conceptually supportive of cognitive flow but practically fragile. While physical integration and volitional activation can enhance agency and reduce screen dependence, they also introduce new forms of cognitive and physical effort when sensing and system feedback are unreliable.

5.2 Legibility, Social Comfort, and the Limits of Episodic Interaction

Participants expressed low comfort using the device in shared settings, primarily due to concerns about audibility and public perception of wearable cameras. At the same time, participants expressed interest in more proactive capabilities, revealing an unresolved tension between intentional control and seamless intelligence. This suggests that episodic interaction does not eliminate the challenges of wearable sensing, but reframes them as balancing the tradeoffs between the utility of always-on systems and the comfort and control provided by episodic boundaries.

5.3 Future Work

Future work should study longer-term use across different settings, including how session boundaries affect personal habits. Additional research is needed on bystander interpretation and social signaling, and perception of alternative boundary-setting rituals beyond wearing (e.g., spatial “zones” and wake word activation). The prototype can also be improved by detecting when the user is wearing the device and providing haptic feedback. Finally, future systems may explore how assistants can better adapt to different types of task episodes while preserving user control and legibility.

Episodic wearable assistance may be particularly meaningful for people with visual impairments. Prior work such as LLM-Glasses demonstrates how AI-driven wearables can support navigation and situational awareness through continuous multimodal feedback [23]. Our findings suggest an alternative paradigm in which visually impaired users intentionally invoke assistance only when needed (e.g., during navigation, shopping, etc.), balancing accessibility with autonomy rather than defaulting to always-on support.

6 Conclusion

We introduced The Pen and demonstrated how episodic embodied interaction reframes wearable AI as an intentional cognitive tool and argued for this paradigm as a default over continuous infrastructure.

Acknowledgments

We thank our study participants for their time and feedback. We also thank our advisor Dr. Nirupam Roy for guidance and support throughout this project.

References

- [1] Rawan Alharbi et al. 2018. I Can't Be Myself: Effects of Wearable Cameras on the Capture and Sharing of Personal Experiences. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018).
- [2] Ryo Arakawa, Hiromu Yakura, and Mayank Goel. 2024. PrISM-Observer: Intervention Agent to Support Everyday Procedures Using Smartwatches. In *Proceedings of the 2024 ACM Symposium on User Interface Software and Technology*.
- [3] Laura Bajorunaite, Julie R. Williamson, and Stephen Brewster. 2024. VR Headsets In-The-Wild: Qualitative Insights on Safety, Awareness, and Social Challenges from Real Train Journeys. In *Proceedings of the 23rd International Conference on Mobile and Ubiquitous Multimedia (MUM '24)*. Association for Computing Machinery, New York, NY, USA, 82–94. doi:10.1145/3701571.3701576
- [4] Victoria Bellotti and Keith Edwards. 2001. Intelligibility and Accountability: Human Considerations in Context-Aware Systems. *Human-Computer Interaction* 16, 2 (2001), 193–212.
- [5] Disha Bhardwaj et al. 2024. In Focus, Out of Privacy: Wearers' Perspectives on Camera Glasses. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. doi:10.1145/3701571.3701578
- [6] Tamara Denning, Zakariya Dehlawi, and Tadayoshi Kohno. 2014. In Situ with Bystanders of Augmented Reality Glasses. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*.
- [7] Cathy Mengying Fang, Yasith Samaradivakara, Pattie Maes, and Suranga Nanayakkara. 2025. Mirai: A Wearable Proactive AI "Inner-Voice" for Contextual Nudging. arXiv:2502.02370 [cs.HC] <https://arxiv.org/abs/2502.02370>
- [8] Changshuo Hu, Qiang Yang, Yang Liu, Tobias Röddiger, Kayla-Jade Butkow, Mathias Ciliberto, Adam Luke Pullin, Jake Stuchbury-Wass, Mahbub Hassan, Cecilia Mascolo, and Dong Ma. 2025. A Survey of Earable Technology: Trends, Tools, and the Road Ahead. (2025). arXiv:2506.05720 [cs.HC] <https://arxiv.org/abs/2506.05720>
- [9] Fang Hu, Peng He, Songlin Xu, Yin Li, and Cheng Zhang. 2020. FingerTrak: Continuous 3D Hand Pose Tracking by Deep Learning Hand Silhouettes Captured by Miniature Thermal Cameras on Wrist. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 2, Article 71 (June 2020), 24 pages. doi:10.1145/3397306
- [10] Jingyang Hu, Hongbo Jiang, Daibo Liu, Zhu Xiao, Qibo Zhang, Jiangchuan Liu, and Schahram Dustdar. 2024. Combining IMU With Acoustics for Head Motion Tracking Leveraging Wireless Earphone. *IEEE Transactions on Mobile Computing* 23, 6 (2024), 6835–6849. doi:10.1109/TMC.2023.3325826
- [11] Umar Iqbal and Mark A. Campbell. 2022. Adopting Smart Glasses Responsibly: Potential Benefits, Ethical Concerns, and Risk Mitigation Strategies. *AI and Ethics* (2022). doi:10.1007/s43681-022-00155-7
- [12] Arnav Kapur, Shreyas Kapur, and Pattie Maes. 2018. AlterEgo: A Personalized Wearable Silent Speech Interface. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces*. doi:10.1145/3172944.3172977
- [13] Maruchi Kim, Antonio Glenn, Bandhav Veluri, Yunseo Lee, Eyoel Gebre, Aditya Bagaria, Shwetak Patel, and Shyamnath Gollakota. 2024. IRIS: Wireless ring for vision-based smart home interaction. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24)*. ACM, 1–16. doi:10.1145/3654777.3676327
- [14] Marion Koelle, Wilko Heuten, and Susanne Boll. 2019. Evaluating a Wearable Camera's Social Acceptability In-the-Wild. In *Extended Abstracts of the SIGCHI Conference on Human Factors in Computing Systems*.
- [15] Mohammed Korayem, Robert Templeman, Dennis Chen, David Crandall, and Apu Kapadia. 2014. ScreenAvoider: Protecting Computer Screens from Ubiquitous Cameras. *arXiv preprint arXiv:1412.0008* (2014).
- [16] Gunhyuk Lee, Meng Xia, Deyi Li, Yuntao Chen, and Ruochen Du. 2025. Sensible Agent: Unobtrusive Interaction with Proactive AR Agents. In *Proceedings of the 2025 ACM Symposium on User Interface Software and Technology*.
- [17] Tianjian Liu, Hongzheng Zhao, Yuheng Liu, Xingbo Wang, and Zhenhui Peng. 2024. ComPeer: A Generative Conversational Agent for Proactive Peer Support. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 117, 22 pages. doi:10.1145/3654777.3676430
- [18] Anna Orchard et al. 2022. Augmented Reality Smart Glasses in Focus: A User Group Study of Emerging Social and Ethical Concerns. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW (2022).
- [19] Tobias Röddiger, Christopher Clarke, and Hans Gellersen. 2021. EarRumble: Discreet Hands- and Eyes-Free Input by Voluntary Tensor Tympani Muscle Contraction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*.
- [20] Tobias Röddiger, Christopher Clarke, Hans Gellersen, and Michael Beigl. 2022. Sensing with Earables: A Systematic Review and Taxonomy. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 1 (2022).
- [21] Jiayu Shu, Rui Zheng, and Pan Hui. 2016. Cardea: Context-Aware Visual Privacy Protection from Pervasive Cameras. arXiv:1610.00889 [cs.CR] <https://arxiv.org/abs/1610.00889>
- [22] Desney Tan, Dan Morris, and T. Scott Saponas. 2010. Interfaces on the go. *XRDS* 16, 4 (June 2010), 30–34. doi:10.1145/1764848.1764856
- [23] Issatay Tokmurziyev, Miguel Altamirano Cabrera, Muhammad Haris Khan, Yara Mahmoud, and Dzmitry Tsetserukou. 2026. LLM-Glasses: GenAI-driven Glasses with Haptic Feedback for Navigation of Visually Impaired People. arXiv:2503.16475 [cs.HC] <https://arxiv.org/abs/2503.16475>
- [24] Yonatan Tussa, Andy Heredia, and Nirupam Roy. 2025. Lessons Learned from Developing a Privacy-Preserving Multimodal Wearable for Local Voice-and-Vision Inference. arXiv:2511.11811 [cs.HC] <https://arxiv.org/abs/2511.11811>
- [25] Wazeer Deen Zulfikar, Samantha W. T. Chan, and Pattie Maes. 2024. Memoro: Using Large Language Models to Realize a Concise Interface for Real-Time Memory Augmentation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3613904.3642450