

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

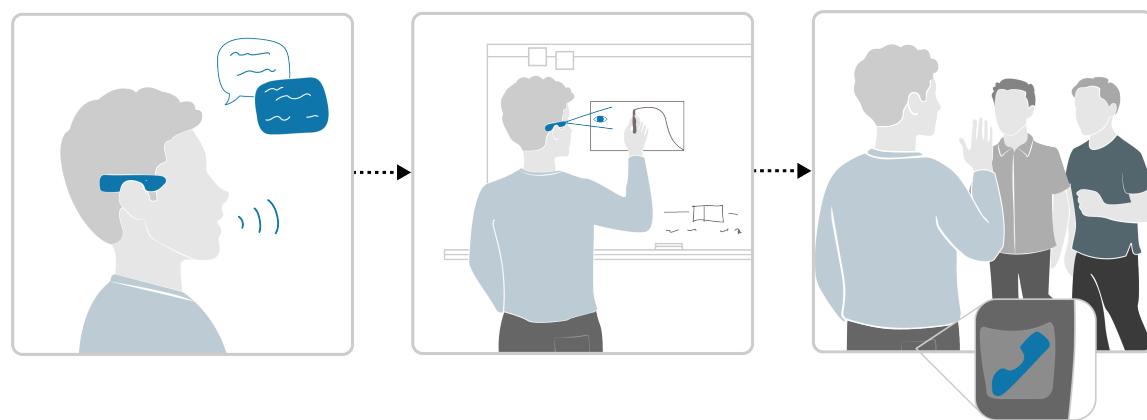


Fig. 1. The Pen is worn behind the ear and the user is speaking to it (left); user continues whiteboarding while the device observes visual context (middle); user disengages by removing the device and placing it in a pocket before resuming social interaction (right).

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 XXXXXXXX>

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

53 **1 Introduction**

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

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

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

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

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

75 *We ask the following questions:*

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

79 *Our contributions are the following:*

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

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

105 2 Related Work

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

112 113 2.1 Ear-Worn and Screenless Wearable Interaction

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

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

127 128 2.2 Situated Assistance and Memory Support

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

136 137 2.3 Activation Tax and Micro-Interaction Costs

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

146 147 2.4 Privacy, Social Acceptability, and Legible Boundaries

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

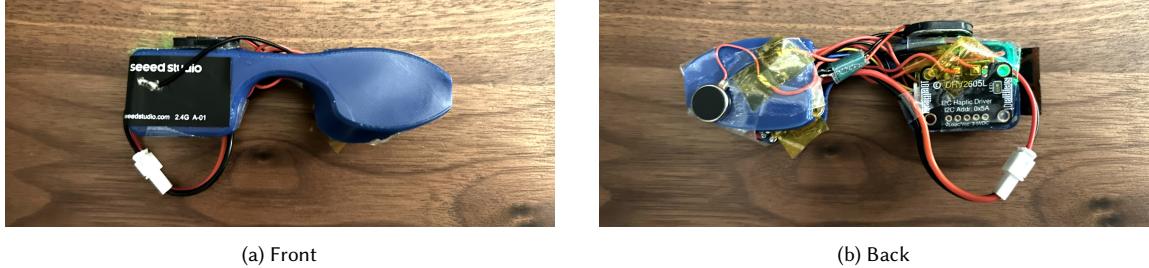


Fig. 2. Front and back of The Pen prototype.

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

209 user understanding of system state. In contrast, removing the device was consistently perceived as a clear end to the
210 interaction: taking it off felt definitive and relieved any sense of ongoing system involvement.
211

212 Participants also experienced episodic use as increasing their sense of agency and intentional control over the system.
213 Several preferred using the device only at selected moments rather than continuously, describing it as something
214 that felt “extra” rather than always necessary. Voice interaction was perceived as natural and expressive, enabling
215 participants to articulate questions without shifting attention to a phone and reducing reliance on manual interaction.
216 Non-native English speakers perceived particular value in voice interaction, noting that handheld interfaces often
217 introduce friction when interacting with systems designed primarily in a non-native language.
218

219 However, this sense of control came with interaction overhead. Participants reported friction when invoking the
220 device, especially when visual capture required repeated attempts. Technical limitations occasionally disrupted task
221 flow, with sensing failures forcing participants to pause their activity and reducing their ability to remain focused in
222 some cases. As a result, the device was perceived as conceptually supportive of cognitive flow but practically fragile.
223 While physical integration and volitional activation can enhance agency and reduce screen dependence, they also
224 introduce new forms of cognitive and physical effort when sensing and system feedback are unreliable.
225

227 **5.2 Legibility, Social Comfort, and the Limits of Episodic Interaction**

228 Participants expressed low comfort using the device in shared settings, primarily due to concerns about audibility and
229 public perception of wearable cameras. At the same time, participants expressed interest in more proactive capabilities,
230 revealing an unresolved tension between intentional control and seamless intelligence. This suggests that episodic
231 interaction does not eliminate the challenges of wearable sensing, but reframes them as balancing the tradeoffs between
232 the utility of always-on systems and the comfort and control provided by episodic boundaries.
233

236 **5.3 Future Work**

237 Future work should study longer-term use across different settings, including how session boundaries affect personal
238 habits. Additional research is needed on bystander interpretation and social signaling, and perception of alternative
239 boundary-setting rituals beyond wearing (e.g., spatial “zones” and wake word activation). The prototype can also be
240 improved by detecting when the user is wearing the device and providing haptic feedback. Finally, future systems may
241 explore how assistants can better adapt to different types of task episodes while preserving user control and legibility.
242

243 Episodic wearable assistance may be particularly meaningful for people with visual impairments. Prior work such
244 as LLM-Glasses demonstrates how AI-driven wearables can support navigation and situational awareness through
245 continuous multimodal feedback [23]. Our findings suggest an alternative paradigm in which visually impaired users
246 intentionally invoke assistance only when needed (e.g., during navigation, shopping, etc.), balancing accessibility with
247 autonomy rather than defaulting to always-on support.
248

251 **6 Conclusion**

252 We introduced The Pen and demonstrated how episodic embodied interaction reframes wearable AI as an intentional
253 cognitive tool and argued for this paradigm as a default over continuous infrastructure.
254

256 **Acknowledgments**

257 We thank our study participants for their time and feedback. We also thank our advisor Dr. Nirupam Roy for guidance
258 and support throughout this project.
259

261 References

262 [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*
 263 *ACM on Human-Computer Interaction* 2, CSCW (2018).

264 [2] Ryo Arakawa, Hiromu Yakra, and Mayank Goel. 2024. PriSM-Observer: Intervention Agent to Support Everyday Procedures Using Smartwatches.
 265 In *Proceedings of the 2024 ACM Symposium on User Interface Software and Technology*.

266 [3] Laura Bajorunaite, Julie R. Williamson, and Stephen Brewster. 2024. VR Headsets In-The-Wild: Qualitative Insights on Safety, Awareness, and
 267 Social Challenges from Real Train Journeys. In *Proceedings of the 23rd International Conference on Mobile and Ubiquitous Multimedia (MUM '24)*.
 268 Association for Computing Machinery, New York, NY, USA, 82–94. doi:10.1145/3701571.3701576

269 [4] Victoria Bellotti and Keith Edwards. 2001. Intelligibility and Accountability: Human Considerations in Context-Aware Systems. *Human-Computer*
 270 *Interaction* 16, 2 (2001), 193–212.

271 [5] Disha Bhardwaj et al. 2024. In Focus, Out of Privacy: Wearers' Perspectives on Camera Glasses. In *Proceedings of the SIGCHI Conference on Human*
 272 *Factors in Computing Systems*. doi:10.1145/3701571.3701578

273 [6] Tamara Denning, Zakariya Dehlawi, and Tadayoshi Kohno. 2014. In Situ with Bystanders of Augmented Reality Glasses. In *Proceedings of the*
 274 *SIGCHI Conference on Human Factors in Computing Systems*.

275 [7] Cathy Mengying Fang, Yasith Samaradivakara, Pattie Maes, and Suranga Nanayakkara. 2025. Mirai: A Wearable Proactive AI "Inner-Voice" for
 276 Contextual Nudging. arXiv:2502.02370 [cs.HC] <https://arxiv.org/abs/2502.02370>

277 [8] Changshuo Hu, Qiang Yang, Yang Liu, Tobias Röddiger, Kayla-Jade Butkow, Mathias Ciliberto, Adam Luke Pullin, Jake Stuchbury-Wass, Mahbub
 278 Hassan, Cecilia Mascolo, and Dong Ma. 2025. A Survey of Earable Technology: Trends, Tools, and the Road Ahead. (2025). arXiv:2506.05720 [cs.HC]
 279 <https://arxiv.org/abs/2506.05720>

280 [9] Fang Hu, Peng He, Songlin Xu, Yin Li, and Cheng Zhang. 2020. FingerTrak: Continuous 3D Hand Pose Tracking by Deep Learning Hand Silhouettes
 281 Captured by Miniature Thermal Cameras on Wrist. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 2, Article 71 (June 2020), 24 pages.
 282 doi:10.1145/3397306

283 [10] Jingyang Hu, Hongbo Jiang, Daibo Liu, Zhu Xiao, Qibo Zhang, Jiangchuan Liu, and Schahram Dustdar. 2024. Combining IMU With Acoustics for
 284 Head Motion Tracking Leveraging Wireless Earphone. *IEEE Transactions on Mobile Computing* 23, 6 (2024), 6835–6849. doi:10.1109/TMC.2023.3325826

285 [11] Umar Iqbal and Mark A. Campbell. 2022. Adopting Smart Glasses Responsibly: Potential Benefits, Ethical Concerns, and Risk Mitigation Strategies.
 286 *AI and Ethics* (2022). doi:10.1007/s43681-022-00155-7

287 [12] Arnav Kapur, Shreyas Kapur, and Pattie Maes. 2018. AlterEgo: A Personalized Wearable Silent Speech Interface. In *Proceedings of the 23rd International*
 288 *Conference on Intelligent User Interfaces*. doi:10.1145/3172944.3172977

289 [13] Maruchi Kim, Antonio Glenn, Bandhav Veluri, Yunseo Lee, Eyoel Gebre, Aditya Bagaria, Shwetak Patel, and Shyamnath Gollakota. 2024. IRIS:
 290 Wireless ring for vision-based smart home interaction. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*
 291 (*UIST '24*). ACM, 1–16. doi:10.1145/3654777.3676327

292 [14] Marion Koelle, Wilko Heuten, and Susanne Boll. 2019. Evaluating a Wearable Camera's Social Acceptability In-the-Wild. In *Extended Abstracts of*
 293 *the SIGCHI Conference on Human Factors in Computing Systems*.

294 [15] Mohammed Korayem, Robert Templeman, Dennis Chen, David Crandall, and Apu Kapadia. 2014. ScreenAvoider: Protecting Computer Screens
 295 from Ubiquitous Cameras. *arXiv preprint arXiv:1412.0008* (2014).

296 [16] Gunhyuk Lee, Meng Xia, Deyi Li, Yuntao Chen, and Ruochen Du. 2025. Sensible Agent: Unobtrusive Interaction with Proactive AR Agents. In
 297 *Proceedings of the 2025 ACM Symposium on User Interface Software and Technology*.

298 [17] Tianjian Liu, Hongzheng Zhao, Yuheng Liu, Xingbo Wang, and Zhenhui Peng. 2024. ComPeer: A Generative Conversational Agent for Proactive
 299 Peer Support. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24)*. Association for Computing
 300 Machinery, New York, NY, USA, Article 117, 22 pages. doi:10.1145/3654777.3676430

301 [18] Anna Orchard et al. 2022. Augmented Reality Smart Glasses in Focus: A User Group Study of Emerging Social and Ethical Concerns. *Proceedings of*
 302 *the ACM on Human-Computer Interaction* 6, CSCW (2022).

303 [19] Tobias Röddiger, Christopher Clarke, and Hans Gellersen. 2021. EarRumble: Discreet Hands- and Eyes-Free Input by Voluntary Tensor Tympani
 304 Muscle Contraction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*.

305 [20] Tobias Röddiger, Christopher Clarke, Hans Gellersen, and Michael Beigl. 2022. Sensing with Earables: A Systematic Review and Taxonomy.
 306 *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 1 (2022).

307 [21] Jiayu Shu, Rui Zheng, and Pan Hui. 2016. Cardea: Context-Aware Visual Privacy Protection from Pervasive Cameras. arXiv:1610.00889 [cs.CR]
 308 <https://arxiv.org/abs/1610.00889>

309 [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

310 [23] Issatay Tokmurziyev, Miguel Altamirano Cabrera, Muhammad Haris Khan, Yara Mahmoud, and Dzmitry Tsetserukou. 2026. LLM-Glasses:
 311 GenAI-driven Glasses with Haptic Feedback for Navigation of Visually Impaired People. arXiv:2503.16475 [cs.HC] <https://arxiv.org/abs/2503.16475>

312 [24] Yonatan Tussa, Andy Heredia, and Nirupam Roy. 2025. Lessons Learned from Developing a Privacy-Preserving Multimodal Wearable for Local
 313 Voice-and-Vision Inference. arXiv:2511.11811 [cs.HC] <https://arxiv.org/abs/2511.11811>

314 [25] Wazeer Deen Zulfiqar, Samantha W. T. Chan, and Pattie Maes. 2024. Memoro: Using Large Language Models to Realize a Concise Interface for
 315 Real-Time Memory Augmentation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3613904.3642450