Powering for Privacy: Improving User Trust in Smart Speaker Microphones with Intentional Powering and Perceptible Assurance

Youngwook Do and Nivedita Arora, *Georgia Institute of Technology;*
Ali Mirzazadeh and Injoo Moon, *Georgia Institute of Technology and Massachusetts Institute of Technology;* Eryue Xu, *Georgia Institute of Technology;*
Zhihan Zhang, *Georgia Institute of Technology and University of Washington;*
Gregory D. Abowd, *Georgia Institute of Technology and Northeastern University;*
Sauvik Das, *Georgia Institute of Technology and Carnegie Mellon University*

https://www.usenix.org/conference/usenixsecurity23/presentation/do

This paper is included in the Proceedings of the 32nd USENIX Security Symposium.

August 9–11, 2023 • Anaheim, CA, USA

978-1-939133-37-3

Open access to the Proceedings of the 32nd USENIX Security Symposium is sponsored by USENIX.

Powering for Privacy: Improving User Trust in Smart Speaker Microphones with Intentional Powering and Perceptible Assurance

Youngwook Do[1], Nivedita Arora[1], Ali Mirzazadeh[*1,4], Injoo Moon[*1,4], Eryue Xu[*1], Zhihan Zhang[1,5], Gregory D. Abowd [1,2], Sauvik Das[1,3]

[1]Georgia Institute of Technology, [2]Northeastern University, [3]Carnegie Mellon University, [4]Massachusetts Institute of Technology, [5]University of Washington

Abstract

Smart speakers come with always-on microphones to facilitate voice-based interaction. To address user privacy concerns, existing devices come with a number of privacy features: e.g., mute buttons and local trigger-word detection modules. But it is difficult for users to trust that these manufacturer-provided privacy features actually work given that there is a misalignment of incentives: Google, Meta, and Amazon benefit from collecting personal data and users know it. What's needed is *perceptible assurance* — privacy features that users can, through physical perception, verify actually work. To that end, we introduce, implement, and evaluate the idea of "intentionally-powered" microphones to provide users with perceptible assurance of privacy with smart speakers. We employed an iterative-design process to develop *Candid Mic*, a battery-free, wireless microphone that can only be powered by harvesting energy from intentional user interactions. Moreover, users can visually inspect the (dis)connection between the energy harvesting module and the microphone. Through a within-subjects experiment, we found that *Candid Mic* provides users with perceptible assurance about whether the microphone is capturing audio or not, and improves user trust in using smart speakers relative to mute button interfaces.

1 Introduction

Smart speakers enable convenient, hands-free interaction with Internet of Things (IoT) devices and mobile applications but raise significant end-user privacy concerns. Prior work has shown that people are concerned that their private conversations could be put at risk of being recorded without their knowledge or consent [27, 54]. One report suggests that more than 50 % of U.S. smart speaker users harbor privacy concerns [35]. A Statista report suggested that privacy concerns are one of the main factors that inhibit smart speaker adoption [45].

In response, smart speaker manufacturers include a number of privacy-enhancing features such as local trigger-word detection modules and mute buttons. These features, however, have done little to allay user concerns. Indeed, though many smart speakers are designed to only be activated by wake-up words (e.g., "OK Google"), they sometimes activate in error even if users do not speak the wake-up word [18, 25, 41, 50]. End-users have noticed these errors, further raising privacy concerns [27]. Moreover, while many smart speakers are equipped with mute buttons that are meant to disable the microphone, users must blindly trust that the microphone is, in fact, muted [ibid]. Overcoming this trust barrier can be a challenge since there is a conflict of interest between many smart speaker manufacturers who monetize personal data (e.g., Meta, Google, Amazon) and end-users who wish to guard against unwitting personal data collection.

To that end, the fundamental question that drove our work was: How can we design smart speaker microphones that users trust are only activated when intended? We hypothesize that doing so should help alleviate users' privacy concerns. To address this question, we employed an iterative, human-centered design process that spanned three phases of work.

In the first phase, we conducted a formative semi-structured interview study with 10 participants. We interviewed 10 participants who did not use a smart speaker system due to privacy concerns and asked them questions to better understand their key concerns and how we might design alternatives that address those concerns from an interaction design perspective. We found that participants wanted to be able to audit and visibly confirm for themselves that a smart speaker microphone was deactivated; for example, they had trouble trusting mute buttons because there remained the possibility that the underlying software could be manipulated by attackers or by manufacturers.

Since many end-users do not trust smart speaker manufacturers to be good data stewards, users wanted *perceptible assurance* that always-on smart speaker microphones are only activated when intended: i.e., to be able to verify, not blindly trust. Perceptible assurance requires direct observation, from

[*]Three authors contributed equally to this research.

users, of whether the sensor is powered or otherwise capable of capture. Mute buttons that toggle software state, or even internal hardware kill switches that disable recording when not in use [20], are invisible to end-users and thus do not provide this perceptible assurance. Prior work examining the privacy behaviors of smart speaker users found that one-way users feel perceptible assurance that their smart speakers are not listening is by disconnecting the microphone from its power source [1, 10, 23, 27, 42]. Other prior work illustrates the importance of physical intuition and inspectability in engendering user trust with camera covers [17]. In short, internal hardware kill switches will not provide users with perceptible assurance without physically intuitive design that users can directly observe. Thus, the central hypothesis of our work is that if we can create *intentionally-powered* microphones that can only be powered through user interactions in a manner that users can physically verify, we should be able to provide perceptible assurance that smart speaker microphones are only activated as intended, and thereby increase user trust.

Building on these insights, in the second phase of work, we designed and implemented *Candid Mic*, a new microphone that provides physically perceivable, verifiable guarantees that the microphone is activated or deactivated in line with user expectations. To do so, we adapted the *MARS* battery-free, wireless microphone [4]. The *MARS* microphone in its original conception harvests energy from an ambient energy source (e.g., ambient light) to employ radio-frequency backscatter to wirelessly transmit an audio signal to an edge-computing hub. We modified the *MARS* microphone in a number of ways: we placed it in a portable clam-shell casing; we replaced the energy harvesting module with a series of photodiodes so it would be powered by ambient light only when the case was opened; and, we created a hinge-apparatus and circuit design that would physically disconnect the power lines between the photodiodes and the microphone so that there would be no possibility for the microphone and the wireless communication module to be powered when the case is closed.

In the third phase of our work, we tested our hypothesis with a summative user evaluation. We recruited 16 participants who have privacy concerns with smart speakers to assess if, and how, the interaction design of our new microphone might address their privacy concerns with smart speaker microphones. We conducted a within-subjects experiment where recruited participants used both a baseline mute-button interface for an Amazon Echo Dot and *Candid Mic* to interact with a smart speaker. The conditions were presented to participants in counterbalanced order. Across both conditions, participants were asked to rate how much they trusted that the microphone was actually muted when they intended it to be muted.

Our findings confirm our central hypothesis that intentional powering and perceptible assurance increase user trust in smart speaker microphones. Specifically, we found that participants trusted the mute operation of *Candid Mic* significantly more than the mute-button interface of the Amazon Echo Dot. Moreover, in explaining their rationale for their ratings in an exit interview, we found that participants' increased trust in *Candid Mic* was due to the fact that they could manually verify that there was a visible power disconnection between the energy harvesting module and the microphone when *Candid Mic* was placed in the mute position.

In sum, we provide three concrete contributions:

- We conducted a formative study and analyzed findings to derive key design considerations that help increase user trust in using smart speaker microphones.
- Building on these design considerations, we implemented and evaluated *Candid Mic*, an "intentionally-powered" smart speaker microphone that provides users with perceptible assurance that, when muted, the microphone is disconnected from its power source and cannot capture audio.
- Through a within-subjects experiment, we show that users trusted *Candid Mic* significantly more than they trusted an Amazon Echo Dot to not be capturing audio when muted. Through additional qualitative analysis, we show that this increase in trust appears to be because of intentional powering and perceptible assurance.

2 Background and Related Work

In this section, we review prior art exploring how current smart speaker designs raise people's privacy concerns. Then, we summarize people's decisions against such privacy concerns and how prior work aims to address those concerns.

2.1 Privacy Concerns against Smart Speaker Eavesdropping

Addressing user privacy concerns with smart speakers is a long-standing problem. Prior work has shown that users are concerned about their smart speakers eavesdropping on their conversations without their knowledge and consent [27, 38]. These concerns remain even in the presence of privacy features such as local trigger-word detection modules: indeed, because these modules are prone to error, they can sometimes falsely activate. When noticed by end-users, these false activations can exacerbate concerns that their private conversations are recorded without their knowledge or consent [27]. Unsurprisingly, privacy concerns like these are one of the main inhibitors of widespread smart speaker adoption [45]. In our paper, we study how to design and develop a device for end-users to ensure smart speaker microphones activate only in alignment with user intentions.

Prior work suggests that there are many factors that influence end-user privacy concerns with smart speakers. For example, end-users harbor privacy concerns about data stewardship after their voice data has already been collected (e.g.,

if there is a data breach) [13]. Additionally, as smart speaker devices are often shared, prior work has also shown how end-users are sometimes concerned about how their private information may be accessible to others who use the smart speaker [22, 29]. Both of these cases imply that audio data has already been recorded and transmitted. In contrast, our focus is on understanding and minimizing privacy concerns related to audio *capture* rather than how the captured data are managed, used, and made accessible to others once captured.

2.2 Solutions to Address Privacy Concerns against Smart Speakers

Prior work has studied and proposed various solutions to address privacy concerns related to unauthorized smart speaker microphone access.

2.2.1 Privacy-seeking Behaviors against Smart Speakers

Smart speaker users have developed their own ad-hoc practices to address their privacy concerns [10, 23, 27]. Even with the knowledge that smart speakers come with mute buttons, end-users often do not trust these software-based mute controls because they believe that these controls can be manipulated [2, 27, 52]. Accordingly, prior work has shown that end-users employ physical and visible tactics to thwart unintended recording. For example, end-users have started unplugging the smart speaker to shut down its capabilities [10, 23, 27, 42]. This physical decoupling of a device from its energy source engenders trust — users have a more physically intuitive understanding of the connection between their actions and the device's function. While unplugging a smart speaker can guarantee that it is deactivated, it also reduces the utility of the device in practice as users will have to plug it in again to access its functionality.

In addition to unplugging, prior work also suggests that end-users cover their smart speakers with physical materials (e.g., towel) and stay away from a smart speaker when they wish not to be heard [2, 37]. However, since sound propagates through many physical barriers, it is unclear to end-users if and to what extent their smart speakers can still "hear" them.

We build upon and address the limitations of these ad-hoc practices by designing and developing a smart speaker microphone that affords end-users physically verifiable clarity on when it is activated and that still allows users to unlock the utility of their devices at will.

2.2.2 Technical Solutions against Unauthorized Microphone Recording

Researchers proposed various ways to thwart covert eavesdropping such as utilizing masking noise, jamming their smart-speaker microphones and providing mechanisms for on-demand power activation. Tung et al. presented a technique of adding masking noise to a user's speech and removing the noise from a receiver end [49]. Roy et al. presented a technique to use inaudible sound to jam an unauthorized microphone recording [39]. Chen et al. introduced a wrist-worn wearable device that allows end-users to emit human inaudible noise that deafens nearby microphones [12]. Additionally, Sun et al. demonstrated a smart speaker accessory device that generates continuous jamming sound to a smart speaker until a user speaks a smart speaker's wake-up word [46]. Similarly, Project Alias demonstrates a way for a user to wake up a smart speaker with their customized wake-up word, jamming the smart speaker when not in use [24]. Chandrasekaran et al. studied two probes that prevent unauthorized smart speaker recording: using a remote-controller-based jamming device and outlet to power on and off a smart speaker [10]. The jamming technique is expected to drown out other sounds in an end-user's private space, but prior work has also shown that it may still be possible to capture private conversation even in the presence of jamming depending on microphone types or jamming sound directions [12, 51]. In addition, while the power on/off solution can guarantee the microphone's deactivation, rebooting the device reduces a smart speaker's usability as end-users have to manually reconnect their devices to power and then wait for the device to be fully boot. As prior work has shown, users have a low tolerance for delays when it comes to security and privacy [19].

As trigger-word detection is prone to accidental activation, Mhaidli et al. investigated alternative activation triggers (e.g., loud voice and gaze) [30]. While such 'wake-up' interactions could prevent a smart speaker's accidental activation, they do not address the root concern that people do not trust smart speaker microphones to be good data stewards and must blindly trust that their devices are not eavesdropping.

Additionally, researchers have studied techniques to improve users' awareness of smart speaker privacy. Mitev et al. demonstrate LeakyPick, a system that detects unexpected data streaming of smart speakers [31]. LeakyPick detects network traffic to allow end-users to understand whether there is unexpected data transmission of smart speakers. In a similar sense, Charyyev et al. present a technique to detect smart speakers' misactivation by analyzing its network traffic [11]. Song et al. studied visual and auditory feedback to help end-users locate their smart devices including smart speakers [43].

In our work, we create a system that both provides physical guarantees against unauthorized microphone recording and provides end-users with perceptible assurance of this guarantee so that they can trust that their smart speaker microphone cannot capture audio when not in explicit use.

2.3 Tangible and Physical Approaches to Transparent Privacy Operations

Privacy controls for sensor-enabled devices are often abstract, software-based, and not fundamentally connected to function

Figure 1: We employed 3-phase study: (i) formative study; (ii) implementation; and (iii) summative study. Through interview study in our formative study, we found design factors that affect the level of privacy concerns about smart speaker microphones. Then, we implemented a working prototype, *Candid Mic*, by modifying *MARS* [4] and integrating the interview findings. Lastly, we assessed if *Candid Mic* helps improve trust in using smart speaker microphones, compared to a commodity smart speaker.

— thus, users find it difficult to discern what data is collected when and by whom [2, 23, 27, 48]. For instance, many laptop webcams come with an associated LED indicator that is designed to communicate when the webcam is active. However, webcams can be activated without turning on their associated LED indicators [7]. Moreover, many end-users are unsure or mistaken about what the LED indicator communicates [2, 36]. Additionally, while a manufacturer claims that a smart speaker has a hardware switch for mute feature [20], this switch is invisible to end-users, which has them blind trust this claim. In short, there is a gulf between how privacy controls for sensor-enabled devices actually work and how people believe they work [2].

In order to address such concerns, researchers have emphasized the importance of making privacy operations more tangible and/or transparent [2, 48, 53]. For example, Do et al. designed and evaluated an intelligent and physical webcam cover that automatically occludes a webcam when it is not in use [17]. Recently, Ahmad et al. found that disconnecting the physical power wire to a microphone could help end-users build trust in the microphone controls of smart speakers through a survey study with virtual prototypes [1].

Building on and extending this prior work, we design and implement a smart speaker microphone that is transparent in its power activation and provides physical guarantees as to when it can and cannot record.

3 Study Structure

As illustrated in Figure 1, we employed an iterative design process consisting of a formative study, implementation, and a summative evaluation. Our formative study comprised a design exploration with interviews to understand design factors that might address eavesdropping concerns with smart speakers Section 4.

Building on our findings from the formative study, we next implemented a working prototype of an intentionally-powered smart speaker microphone — *Candid Mic* — that provides users with perceptible assurance of when it is activated and deactivated. To do so, we adapted and modified *MARS* microphone system [4], which we will discuss in Section 6.2.

Finally, for the summative evaluation, we conducted a within-subjects lab experiment comparing *Candid Mic* to a commodity smart speaker mute-button interface. Our goal was to assess if and why using *Candid Mic* increases the end-user trust that their smart speaker can only "listen" to their voice when intended.

4 Initial Design Consideration

For our formative study, we started by distilling key design considerations for designing a privacy-preserving smart speaker microphone based on end-user privacy concerns identified in prior work.

4.1 On-demand, User-powered Microphone

Prior work suggests that one key privacy concern with smart speakers that end-users have is that the microphone is "always on," such that a user's conversations can always be recorded without their knowledge or consent [27]. To mitigate this concern, prior work has noted that users employ a number of ad-hoc strategies: e.g., unplugging the smart speaker when not in explicit use [1, 10]. In other words, cutting power is one way in which users can regain the trust that their microphones are not "listening." Accordingly, we explore creating a battery-free, user-powered microphone that can only be powered through intentional interactions.

We draw from and build on the *MARS* self-powered microphone [4]. *MARS* wirelessly transfers audio signals without any battery by harvesting energy from ambient light. This means that *MARS* stays deactivated and cannot record or transmit any audio signals if in a dark environment. Thus, a user can illuminate *MARS* to activate its microphone and vice versa.

This example hints at a way we might align microphone activation with the user intention. For example, *MARS* is powered on when there is energy harvested, and powered off if there is no energy harvested. Thus, if it is *only* possible to intentionally harvest energy, the user can guarantee that the microphone is powered off when they have no intention to use it: i.e., without intention, there is no power.

There are a number of ways to intentionally harvest energy. We brainstormed three that may be both usable and appropri-

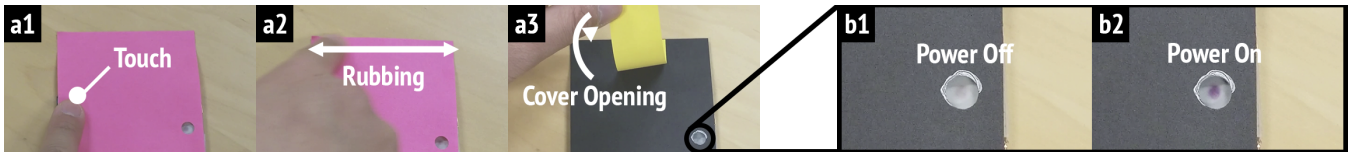

Figure 2: For our formative study, we illustrate various ways to harvest energy to power a microphone via three mock-up prototypes: (a1) touch by body heat; (a2) rubbing by kinetic force; and (a3) opening a cover to expose ambient light to solar cells. Additionally, we also showcase an electrochromic ink display (ECD) as a visual indicator that could be powered by each energy harvesting method. (b1) When no power is applied to the ECD, no image shows up. (b2) Otherwise, a red-dot shape shows up.

ate for the microphone context: (i) touching one's finger to a thermoelectric generator to harvest energy from the finger's heat; (ii) performing a rubbing motion to convert kinetic energy to electric power; and (iii) opening a cover that exposes photovoltaic cells to harvest energy from ambient light.

4.2 Trusted Indicators

Because microphone operations are opaque and abstract, users need trusted indicators to be better informed about when their microphones are active and listening [17,26,43]. For example, as of iOS 14, Apple introduced a small orange dot-shaped indicator to alert users whenever an iOS app access the host devices' microphone [3]. In the Apple example, however, trust is achieved through a separation of interests: *iOS* indicators inform end-users when a *third-party application* is accessing the *user*'s microphone. Since the interests of iOS and Apple are separate from the interests of the third-party application aiming to access the user's microphone, the indicator is more trustworthy. No such separation of interests exists for smart speaker platforms, however, where the device manufacturer is itself one of the key threats. Even in the Apple example, however, the indicator is a superficial addition that is not inherently tethered to functionality but a software-based trigger-action — it remains plausible that the microphone can "listen" even if the indicator is inactive. Because privacy-concerned users are unlikely to give smart speaker manufacturers the benefit of the doubt, we need trusted, transparent operation such that a user can themselves audit that the (de-)activation of a sensor-use indicator is aligned with the (de-)activation of the sensor.

4.3 Mock-up Prototypes

Based on these design considerations, we made three different paper prototypes corresponding to each of the three different energy harvesting methods discussed in Section 4.1 — touching, opening the cover, and rubbing. For all the mock-up prototypes, we included an activation indicator that would be *physically* connected to the power source: i.e., if the microphone had power, the indicator would be on. We next conducted a formative user study to solicit user feedback on which method was most appropriate and amenable.

5 Formative Study

To understand how users might respond to the idea of intentionally powered microphones, as well the perceived trade-offs between the different mock-up prototypes we created, we conducted an interview study with 10 participants who expressed having privacy concerns with smart speakers.

5.1 Method

5.1.1 Recruitment

We used Prolific and Qualtrics to source participants for our study. We pre-screened participants who reported that they do not use smart speakers mainly because of privacy concerns.[1] Specifically, we asked respondents why they chose *not* to use a smart speaker from the following list, presented in random order—(i) privacy concerns; (ii) lack of utility; (iii) expensive price; and (iv) others. We then invited participants who selected privacy concerns as their primary reason for not adopting a smart speaker to semi-structured interviews as they are our target participants. Since privacy looms large for these users, addressing their concerns is likely to address the concerns of those who already use smart speakers but also harbor privacy concerns. Details of the participants of the formative study can be found in Table 1.

5.1.2 Procedure

We recruited 10 participants from the pre-screening survey. We conducted the interviews remotely over a video conference call. For the first part of the interview, we asked participants about their privacy concerns with smart speakers. For the second part, we showed them a short video illustrating our target problem domain: many people express concern about smart speakers recording their conversations without their explicit knowledge or consent. We next clarified that our new smart speaker microphone is designed to address this target problem domain. Then, we showed participants another video clip that described the general concept of *Candid Mic* and three mock-up video prototypes of the different intentional

[1] We provide our pre-screener in the Appendix.

Table 1: Demographics of the formative study participants.

Gender		Age		CS Education		CS Career	
Male	5	18-24	2	CS	6	CS	6
Female	5	25-34	3	Non-CS	4	Non-CS	4
		35-44	4				
		45-54	-				
		55-64	1				

interactions from which *Candid Mic* could harvest energy to power itself as needed and only as needed. The videos portrayed how end-users might interact with the microphone and smart speaker via paper prototypes. Participants were first asked to provide general feedback on each design concept, and were then asked if and how the components of each design helped address their privacy concerns. The order of the three design ideas was counterbalanced to mitigate order effects. The full set of questions for the interview is provided in the Appendix.

5.1.3 Ethics and Compensation.

Our survey and interview study were IRB-approved. Participants completed the pre-screening survey in 1 minute on average and received $0.20 USD ($12/hr) as compensation upon completing the study. In addition, we provided $6 USD ($12/hr) for the completion of the 30-minute long follow-up interview session.

5.1.4 Data Analysis

We used qualitative coding on our interview data and, then, conducted a thematic analysis to understand emergent themes of the qualitative data from the interview [6,28]. Our goal was to uncover design suggestions critical to lowering end-user privacy concerns with smart speaker microphones. After the research team transcribed the interviews, one member generated the codebook for the three intentional, energy-harvesting interactions— touching, opening the cover, and rubbing—as we wanted to examine which components of each interaction type mattered to participants. Another team member independently followed the codebook and coded the data. Then, the two researchers collaboratively identified larger themes of the participants' responses by performing axial coding [15].

5.2 Results

We found three design factors that build end-user trust in using smart speakers: (1) physically perceivable operations; (2) simple interactions for trust and preferences; and, (3) noticeable activation and deactivation.

5.2.1 Physically Perceivable Operations

Many participants pointed out that physically perceivable operations would help build trust in using a smart speaker by expressing concerns about the commodity smart speaker mute button. For example, P6 illustrated the uncertainty of the mute button: "Even if ... I've muted my smart speaker, how do I know it's actually muted?" Also, P4 also echoed this argument by saying "I'm like, aware of that mute button. But ... is that convincing enough?" This resonates with prior work's findings that users have to blindly trust mute features in smart speakers despite their concerns that the mute features could be manipulated [2,27].

To address such concerns, participants pointed out that physical solutions could improve their level of trust in using smart speaker features. Participants mentioned that they would distrust the claims that the manufacturer made unless they saw a physical disconnection. P8 said, "... like physical space in between an electrical connection maybe ... physical switch, that would make me more confident." P9 also shared that they left a smart speaker unplugged due to privacy concerns. Indeed, prior work has shown that some privacy-conscious users manually unplug their smart speakers when they want to have private conversations [10].

Some participants mentioned that they would love to have a physical cover to prevent sound recording by comparing it to a physical cover for cameras. As P1 suggested, "it would need to ... be constructed of proper materials that would prevent sound coming through properly." Though this suggestion is technically challenging as sound can propagate via physical media unless in a vacuum, participants generally emphasized the importance of introducing physical, perceptible, and user-verifiable assurance about their microphone deactivation.

5.2.2 Simple Interactions for Trust and Preferences

Many participants concurred that they would have more trust in smart speaker microphones if they had more knowledge about how the smart speakers' privacy features worked. To that end, participants mentioned that simple device interactions would help increase trust. P2 expressed confidence that simple devices work as expected by saying "...simple electronics do get working properly I imagined." When talking about the cover interaction mock-up prototype, P7 said "I don't know a ton about electronics. But I think the mechanism makes sense." Additionally, participants who had an understanding of energy harvesting methods—body heat, solar cell, and rubbing— expressed their awareness that the microphone would be powered only when energy is harvested via such methods. P8 mentioned "... when I saw that rubbing [interaction] ... it wouldn't be powered unless the user input energy.."

While participants appreciated the understandability of simple interactions for microphone activation, some showed their concerns that some simple energy harvesting interactions could be susceptible to accidental activation. For example, when comparing touch, rubbing, and cover interactions of the mock-up prototypes, several participants mentioned that

the touch button could be accidentally pressed by objects co-located with the new microphone. P9 speculated: "if you set your coffee cup down there by accident ... I was just curious if that would activate it..." In this sense, participants (P4, P7, P9) mentioned that rubbing interactions would help prevent accidental power activation. However, P6 and P7 discussed the trade-off between usability and security for rubbing interactions: while rubbing interactions would likely prevent accidental activation, it would require too much effort from the user to activate when intended.

5.2.3 Noticeability of Microphone Activation and Deactivation

Participants found that they would need a clear indication as to the activation status of the microphone to build trust in using it. For example, participants (P7, P8, P10) mentioned that the power-source connected indicator in our mock prototypes helped them understand the microphone's activation status, which, in turn, engendered trust. Additionally, participants mentioned that the need to open and close the cover in one of our mock prototypes also made the microphone status more noticeable and tangible as compared to touch and rubbing interactions. P2 said, "... the opening shut design ... along with the indicator, ... so there's like two levels of knowing that it's not off or on."

However, several participants were skeptical about the effectiveness of the indicator display, suggesting that it too could be manipulable. For example, P5 said "it could just be a light that turns on when I touch the button, rather than actually turning the microphone on." P1 also mentioned ".. there's far more reasons to be skeptical than to trust just inherently because they say there's a little dot on it that tells me if it's recording or not." These comments further highlight the need for physical, perceptible, and user-verifiable connections between the functional components of the microphone and its power source.

6 *Candid Mic*

6.1 Design Considerations of *Candid Mic*

Building on our findings from the formative study, we implemented an operational prototype we call *Candid Mic*. We selected the clamshell cover design as it was deemed to be low-effort, resilient to accidental activation, and made it easy for users to verify whether or not the microphone was activated. We discuss our design in detail in the following subsections.

6.1.1 Threat Model

Our microphone is designed to protect against an adversary who can remotely access a specific smart speaker without the user's knowledge or consent. This could include, for example, the device manufacturer or a malicious third-party application. The adversary cannot physically access the smart speaker or the smart speaker microphone to modify its hardware after it has been acquired by the end-user. The objective of this adversary is to monitor or surveil a user through the microphone. We do not consider threats that can access the microphone audio once it has been intentionally activated by the end-user.

6.1.2 Physical and Visible Power Source Disconnection for 'Intentionally-powered' Microphones

In our formative study, we found that users distrust microphones when their device operations are not user-visible and verifiable. For example, while Amazon claims that their Amazon Echo smart speakers are disconnected via a hardware switch when muted [20], this internally embedded switch is invisible to end-users and thus does not provide user-verifiable, perceptible assurance. Unsurprisingly, prior work has shown that owing to this distrust, users take to unplug their smart speakers altogether when they want to be assured that their microphone is not listening [10]. This unplugging affords users a verifiable, physical guarantee that their smart speaker microphone can no longer "listen" [1, 14]. In that vein, Ahmad et al. showed that hardware designs that visually illustrate physical wire disconnections provide more transparency than software-based mechanisms [2].

Accordingly, one key design goal was to provide users with a verifiable, physically perceptible disconnection between the microphone and its power source. Doing so allows us to create "intentionally-powered" microphones that provide users with perceptible assurance. To do so, we created a clamshell hinge design where the power source sits on the bottom of the shell, the microphone sits on the top of the shell, and the connection between the two runs through the hinge. When the cover is closed, the power source and the microphone are visibly disconnected; when it is open, they are visibly connected. The touch and rub-based designs we mocked up and evaluated were less conducive to clearly visible power disconnection.

Beyond the visible disconnection between the power source and microphone, we also utilized a power source that harvests ambient energy in a manner that is only physically possible if the microphone cover is open. Specifically, we used a series of photodiodes that cannot harvest energy when the cover is closed, as the cover blocks external light sources. When the cover is opened, however, the photodiodes will be exposed to ambient light and can harvest this energy to enable microphone function. The combination of the clamshell cover design and the use of the photodiodes as a power source makes for an "intentionally-powered" microphone that affords users perceptible assurance of its status as activated or de-activated.

Figure 3: When unmuting *Candid Mic*, (a) a user presses a lever to open the cover of *Candid Mic*, and (b) says voice commands once the cover is open. (c-d) Otherwise, the user can manually close the cover.

6.1.3 Improving Noticeability by Cover Interaction and Indicator

The clamshell design has an additional benefit: making the microphone activation status obvious based on the cover's positioning. When the cover is open, its shape changes drastically relative to when it is closed. In contrast, the touch and rub-based designs do not entail easily visible changes when the microphone transitions between active and inactive modes.

We also added an indicator display to provide visual feedback when the microphone is powered, in line with the basic usability principle of providing users with a clear indication of the underlying system state [33, 34]. To do so, we used a low-energy visual display indicator that is only powered when the power source for the microphone is connected (i.e. when the clamshell is open). In this way, the indicator provides both visual and functional feedback to end-users. We will discuss the implementation details in the following section.

6.2 Implementation

As we discussed in Section 4.1, we built *Candid Mic* based on *MARS* — a self-powered, wireless microphone [4]. We adapted *MARS* in light of the aforementioned design goals and discuss specific implementation details in the following subsections.

6.2.1 Thin Self-powered Microphone and Acoustic Wireless Communication

To provide users with perceptible assurance that *Candid Mic* cannot be powered without intention, we required a self-powered microphone that can both capture and wirelessly transmit audio signals. We began implementation by adapting *SATURN*, a thin self-powered microphone component used in *MARS* [5]. *SATURN* can capture audio signals, and change its own capacitance accordingly. We use these capacitance changes to enable wireless communication via the frequency-modulated backscatter (FM backscatter) technique [4]. While conventional wireless communication (e.g., Bluetooth, Wifi) generally requires more power than the power that can be harvested in everyday settings, FM backscatter allows the audio signals received by *SATURN* to be transmitted wirelessly with the power harvested in everyday environments, such as ambient light. The power harvester will be discussed in the following subsection. Thus, the FM backscatter technique utilizes frequency modulation that stems from an oscillator for which the operating frequency is dependent on the capacitance changes of *SATURN*. The radio frequency receiver receives the modulated frequency and demodulates the signal to acoustic signals.

6.2.2 Power Harvester

For both the display and oscillator power, we used modular photodiodes that can each provide 250nW of power (1uA at 250mV) in ambient light settings. We place multiple photodiodes in parallel and series to scale to the power requirements of the board oscillator. We describe the detailed arrangement of the photodiodes in Section 6.2.4.

6.2.3 Low-power Indicator Display

For the visual indicator, we required a low-power display that could run on the amount of energy that our photodiodes could harvest — that immediately eliminated more typical illuminators like LEDs. Accordingly, we used an electrochromic display (ECD) as a low-power and lightweight indicator for *Candid Mic*. Specifically, we used an electrochromic polymer material called ECP-Magenta. It is a copolymer consisting of an alkoxy-functionalized 3,4-propylenedioxythiophene (ProDOT) copolymerized and a minimally substituted ProDOT unit [21]. ECP-Magenta changes color from magenta by default to clear when 0.45 V voltage is applied. We also placed green paper below the ECD. As a result, the indicator looks magenta when unpowered and green when powered (as the magenta color clears when power is applied) (See Figure 4 (a4) and (b3)).

6.2.4 Case Design and Functional Units Arrangement

We arranged the functional units of *Candid Mic* inside a clam-shaped casing. The case consists of a cover and a bottom layer which are connected via a hinge. We designed the cover to be open when a user presses a lever once on the side of the case (See Figure 3 (a-b)). To do so, we placed a spring at the hinge which helps the cover stay open until the user manually pulls down and closes the cover (See Figure 3 (c-d)).

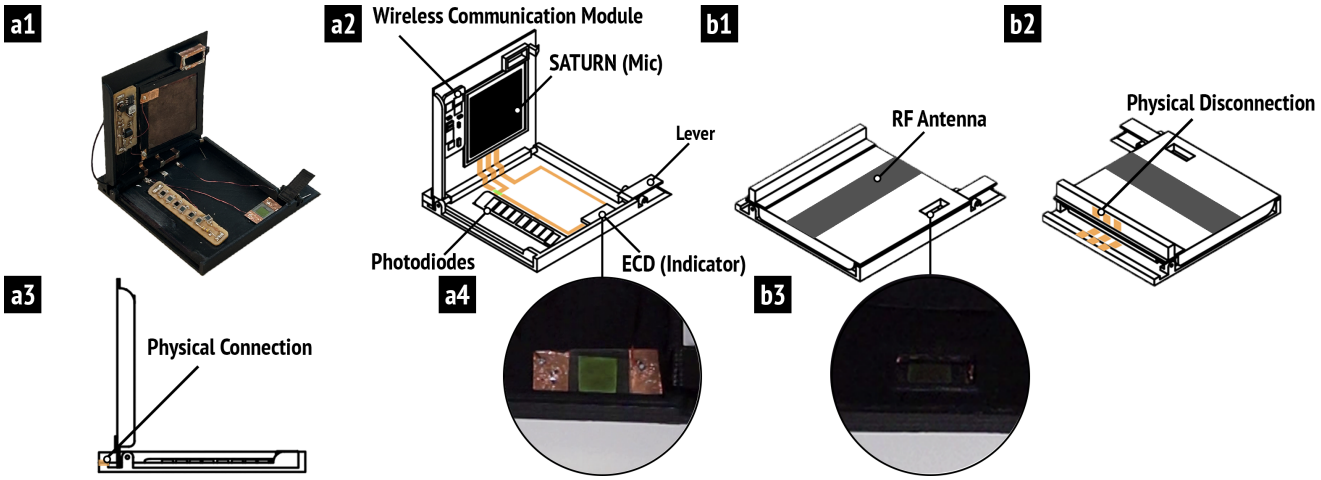

Figure 4: We implemented (a1) *Candid Mic* by modifying *MARS* and integrating findings from our formative study. End-users can unmute and mute the microphone by opening and closing the cover of *Candid Mic*. (a2) End-users can open a cover of *Candid Mic* to unmute a microphone. The wireless communication module, *SATURN* microphone, ECD indicator, and photodiodes are set up inside the cover. (b1) The microphone is muted when the cover is closed. (a3) When *Candid Mic* is open, the circuit wires of all the modules on the hinge of the cover make physical and visible contact, which enables unmuting the microphone. (a4) Simultaneously, the ECD indicator turns transparent and shows its green substrate. (b2) Otherwise, the wires cannot make physical contact, which mutes the microphone. (b3) Then, the ECD indicator turns dark, which indicates the power is drained.

Figure 4 shows how each module is arranged in the casing. The wireless communication module and the microphone (*SATURN*) are placed inside the top lid cover of the casing and the RF antenna is placed on top of the cover(See Figure 4 (a2) and (b1)). The array of photodiodes and the ECD are placed at the bottom cover lid (See Figure 4 (a2)). The photodiodes are arranged in such a way, that, in the presence of light, they produce enough power to light up both ECD and *MARS* . Two photodiodes are in series with a current limiting resistor to be able to operate the ECD indicator at 400mV startup voltage, and five other photodiodes are in parallel to power up the *MARS* circuitry at 200mV. We arranged the photodiodes in this way because of the mismatch in operating voltage between the ECD indicator and the *MARS* circuitry as well as to save on the circuitry cost for power management.

When the box is closed and blocks the light from the photodiodes, the photodiodes cannot harvest power and the microphone is un-powered. Also, we placed a conductive sheet that makes contact with the ECD electrodes when the cover is closed, which allows the ECD display to discharge and turn dark (See Figure 4 (b3)). In short, there is a physical guarantee that the wireless communication is inactive and that the indicator is in-sync with the microphone state.

In addition, the power lines that connect the functional components on the top of the lid to the photodiodes at the bottom run through the hinge of the case (See Figure 4 (a3) and (b2)). Thus, when the cover is open, the physical connection between the top and bottom of the cover is complete, allowing the photodiodes to supply power to the *MARS* circuitry and *SATURN* for active communication. The power from the photodiodes also turns the ECD display clear, allowing users to see the green paper underneath. When the cover is closed, the physical connection between the top and bottom of the case is broken, effectively deactivating the functional components and turning the ECD indicator magenta. In sum, opening and closing the clamshell cover acts as an ON/OFF switch for *Candid Mic*, providing users with perceptible assurance as to the state of the microphone.

7 Summative Study: User Evaluation

Finally, we conducted a summative evaluation to assess if *Candid Mic* — in providing intentional powering and perceptible assurance — increases end-user trust that smart speakers are only listening when the user intends it to be listening. To do so, we conducted a within-subjects experiment comparing *Candid Mic* to a popular commercial baseline.

7.1 Method

7.1.1 Recruitment

We pre-screened participants who were eligible across two categories. First, we recruited people who do not use smart speakers at least partially because of privacy concerns. Second, we recruited people who use smart speakers but have significant privacy concerns about eavesdropping / unintentional listening. We recruited participants from social media

Table 2: Demographics of the summative study participants.

Gender		Age		CS Education		CS Career	
Male	8	18-24	7	CS	12	CS	12
Female	8	25-34	8	Non-CS	4	Non-CS	4
		35-44	-				
		45-54	-				
		55-64	1				

platforms (e.g., Reddit, Slack), email lists, and attached flyers around our institution to look for participants within an accessible range of our office space. Details of the participants of the summative study can be found in Table 2.

7.1.2 Procedure

We ran a within-subjects experiment with two conditions presented to users in counterbalanced order: *Candid Mic* and Amazon Echo Dot. Specifically, our goal was to comparatively evaluate users' trust in the *mute and unmute operations* of *Candid Mic* and Amazon Echo Dot. We chose Amazon Echo Dot for our baseline because it is reported to be the most popular smart speaker in recent years [44].

We first asked participants questions to understand their privacy concerns with smart speakers. Next, we gave participants one of either *Candid Mic* or the Amazon Echo Dot. In both conditions, participants were able to see and manipulate the corresponding system. For *Candid Mic*, participants could directly observe that the top and bottom covers were physically disconnected when closed, as well as observe the ECD indicator state throughout. For the Amazon Echo Dot, participants could press the mute button, which changes color based on its mute state (though the internal mechanics of the mute operation are not visible). We then asked them to watch a video that illustrates how to use the device and how to both mute and unmute the device. After watching the video, we asked them to ask the device about today's weather to familiarize them with how to use a voice command. Then, we let participants freely interact with the devices and specifically asked them to test the device's mute and unmute operations. We next invited participants to offer general feedback on the device they had just used, and then we asked them to fill out a questionnaire on Qualtrics. The questionnaire inquired about trust and usability in using the device's mute and unmute features. Specifically, we operationalized trust as participants' belief as to whether or not the microphone was capable of capturing audio when muted. We avoided directly asking users about "trust" because trust is a broad and multi-faceted concept —affected by confidence, competence, and other factors [47]— and, thus, could be interpreted differently by different people.[2] After participants completed this procedure for the first device, they were given the other device and repeated the same steps. The order in which participants were presented with *Candid Mic* and the Amazon Echo Dot was counterbalanced across participants.

[2]We include our study protocol in the Appendix.

7.1.3 Wizard of Oz Study

The current *Candid Mic* prototype works, but requires a very precise set-up. For example, its performance is sensitive to the orientation of the prototype compared to the RF backscatter antenna. Also, *Candid Mic* requires consistent exposure to ambient light. These requirements made it challenging for end-users to freely interact with *Candid Mic* (e.g., moving it around in a room, etc.). We consider these engineering problems that can be solved now that we have demonstrated a proof-of-concept. As our focus was to evaluate how the design and interaction of *Candid Mic* affect users' trust in using the microphone rather than assessing if the prototype works without error, we employed the Wizard-of-Oz method for our user study [16].

To do so, two researchers ran the study. One researcher ran and moderated the user study by directly interacting with a participant. The other researcher hid behind a one-way mirror in the same room and replicated *Candid Mic* interactions by observing the participant's interaction with *Candid Mic* through the mirror. Specifically, when the participant opened the *Candid Mic* cover and initiated a voice command, their voice was recorded by a hidden microphone and was later fed to the Alexa app. The response of the Alexa app's voice agent was streamed through a wireless speaker placed in front of the participant. If the participant closed the cover, the second researcher stopped feeding the recorded voice to the Amazon app. However, we kept the ECD functional as the ECD indicator could not be remotely replicated by the other researcher. In order to ensure consistent power to the ECD, we replaced the photodiodes with a 1.5V coin cell battery. We covered the battery with an opaque material so that it would be hidden from the participant.

7.1.4 Ethics and Compensation.

Our IRB-approved in-person user study took participants between 24-55 minutes to complete. We debriefed and compensated participants with a $10 USD gift card upon study completion.

7.1.5 Data Analysis

We employed the same qualitative data analysis procedure we performed in Section 5.

7.2 Results

7.2.1 Qualitative Data Analysis

User perceptible power disconnection improves trust in microphone muting. Many participants appreciated the physical, verifiable power disconnection of *Candid Mic* when it was in its mute state. Through this power disconnection,

users felt assured that the microphone was muted. For example, P13 mentioned "it's physically disconnecting the power, gives you more confidence." Of particular note, even participants who self-reported as having little knowledge of electronics understood the implications of *Candid Mic*'s power-cut mechanism. As P2 mentioned: "even like very novice person like me, ... the logic [is] right. ... you know that [if] all circuit is disconnected, then it's not gonna work." Several participants shared that they routinely unplug their smart speakers for privacy reasons and considered *Candid Mic*'s power-cut mechanism as congruent with their desired approach. P12 stated: "I think it's pretty cool how it was explained that it stops the power source to the microphone, because ... how I [mute] the Google Home at my family's home sometimes is by unplugging it physically." In addition, some participants mentioned that the ECD indicator — which gradually runs transparent or opaque when power is or is not applied, respectively — helped them understand when the circuit was connected, disconnected, and in transition. For example, P10 said "[When the circuit is disconnected,] I have the indicator, which is slowly going down. So, I see the power is running off."

However, several participants expressed their concern that *Candid Mic*'s power-cut mechanism could also be deceptive. Since they may not, themselves, know how the circuitry works, participants raised the possibility that the microphone could still potentially be powered in hidden and illicit ways. As P6 stated: "I see the wires disconnected, but I feel that potentially it could be deception." This concern of deception extended to the ECD indicator as well. For example, some participants postulated that the indicator could work independently of the microphone. They did not immediately see nor fully understand how the indicator's and microphone's activation were related. P3 mentioned "if [the indicator display] is connected to other circuits that [are] unrelated to the [microphone] function... then it could be false." While participants showed such concerns, they still appreciated the design attempt to make the operations more transparent to end-users than commodity smart speakers. P3, for example, said: "The reason why I like to see the circuits inside because it's kind of gave me a little bit of sense that I know how it works."

***Candid Mic* is simple, usable, and portable, but should also be durable and sturdy.** While several participants discussed trust as the main reason they would or would not want to adopt *Candid Mic* over extant mute-button interfaces, other participants discussed usability in more depth.

First, many participants (P1, P4, P8, P9, P10, P13, P16) valued portability, with some citing it as a critical factor in determining whether or not they would like to use *Candid Mic* for their smart speaker devices. Owing to the fact that *Candid Mic* is wireless and mobile, their use of and access to their smart speaker would no longer be restricted to the room where it was placed. P1 said "I liked [that] it's wireless... So I guess the idea for this is that it could be like anywhere." Some participants appreciated that the lightweight and compact size of *Candid Mic* amplified its portability. Participants (P4, P8, P13) also appreciated that *Candid Mic* harvested its own energy since it would not require recharging or battery replacements.

Second, most participants found *Candid Mic*'s open-and-shut interaction simple and easy-to-understand. P6, for example, mentioned that a more complicated design might sow doubt about how the microphone works, echoing findings of our formative study in Section 5. Other participants felt that the interaction was not necessarily any simpler than existing mute button interfaces. For instance, P8 said that the lever interaction requires a similar amount of effort as a button press as both need to be manually controlled. P3 added that while there might be a small difference in effort, there can also be a difference in *perceived* effort: "Even though [pressing the lever] requires a bit more labor [than the button press for Amazon Echo], ...cognitively, it seems like that is much more labor than this." We did not find conclusive evidence to suggest that participants viewed either system's manual mute / unmute controls as more or less usable.

Participants expressed concerns about the durability and stability of *Candid Mic* . While they understood that *Candid Mic* is at the prototype phase, they stressed that it would need to be durable and resilient for them to consider its use in practice. P10 said "I don't want [a smart speaker] to be that fragile. So for example, ... you have stuff laying around flying around, it gets underneath a newspaper or something, you wipe it from the table, and it's broken." Additionally, P7 expressed concerns that the wires for the physical connections could get rusty over time if located in a moist room. Also, some participants (P1, P8, P10) experienced a delay in response or no response from *Candid Mic* during the evaluation when they would expect *Candid Mic* to be activated and to respond promptly to their command when in use.

The clam-shell makes it obvious when *Candid Mic* is (un)muted but the ECD should be faster and richer. Many participants discussed how the difference in shape when the *Candid Mic* cover was opened vs. closed provided noticeable visual feedback as to whether *Candid Mic* was muted or not. P13 found that this opening of the cover afforded better visual feedback than the Amazon Echo Dot when the mute button is pressed. As illustrated by P9: "opening this [cover] up means it's unmuted, and then closing [the cover] means it's muted. It's pretty clear. [However] I feel like for smart speakers, there's usually not that much of a visual signal when they're listening." P4 added that this shape change is so noticeable that it might draw too much attention: they would rather want their smart speaker to be blended into the background.

P12 found the ECD indicator helpful in verifying whether the microphone was powered, or not. However, several participants (P1, P12, P14) found the ECD indicator not noticeable

enough. Some highlighted that it was too slow. For example, P12 mentioned: "I really have to focus on one place to see if it's green or not, if it's ready to use. So, that requires a little bit more attention." Additionally, P1 mentioned that because *Candid Mic* does not provide feedback as to whether or not it is capturing acoustic signals, it was difficult to assess whether or not their commands were being heard.

7.2.2 Quantitative Data Analysis

Participants rated the Amazon Echo Dot and *Candid Mic* on how confident they were that, when muted, the microphone was not capturing any audio signals. Fourteen participants rated *Candid Mic* higher than, one participant rated *Candid Mic* equivalent to, and one participant rated *Candid Mic* lower than the Amazon Echo Dot (See Figure 5 (a)). The median ratings for *Candid Mic* and the Amazon Echo Dot were 5 and 2.5, respectively — a substantial difference in favor of *Candid Mic*. To assess the statistical significance of this difference, we performed a Wilcoxon signed-rank test. The result suggests that participants are significantly more confident that *Candid Mic* does not record their conversations when muted than the Amazon Echo Dot ($W=9.5$; $Z=-2.88$; $p<0.01$; $r=0.74$).

We also asked participants to rate both microphones on how confident they were that, when unmuted, the microphone would adequately capture their voice commands. Ten participants rated *Candid Mic* equal to the Amazon Echo Dot while six participants rated *Candid Mic* lower (See Figure 5 (b)). The median ratings for *Candid Mic* and the commodity smart speakers were 4.5 and 5, respectively, suggesting that both microphones were rated highly by all participants. A Wilcoxon signed-rank test confirmed that this difference was statistically significant($W=0$; $Z=2.16$; $p<0.05$; $r=0.88$).

Our qualitative and quantitative analyses suggest that users trust *Candid Mic* significantly more than the Amazon Echo Dot to not record them when they do not want to be recorded. Moreover, the driving force behind this increased trust appears to be intentional powering and perceptible assurance. However, users also expressed concerns about the reliability, durability, and availability of *Candid Mic*, believing that it was slightly less likely than the Amazon Echo Dot to capture the voice they did want to issue to their smart speakers.

8 Discussion

Through our iterative design process, we found that *Candid Mic* reduces end-user privacy concerns related to smart speaker eavesdropping through its novel interaction and design affordances. We now discuss design implications that may help address privacy concerns against smart speaker microphones, remaining challenges in the design of *Candid Mic*, limitations, and opportunities for future research.

8.1 Perceptible Assurance and Intentional Powering Leads to Higher Trust

We found that when users can, for themselves, verify that there is a physical disconnection between a microphone and its power source, and that the microphone can only be powered if they take intentional action, they are more likely to trust that the microphone is only listening in line with their expectations. The clam shell cover design we introduced, where power lines run through the hinge, is one promising, and now tested, way to provide this perceptible assurance. Moreover, the significant shape change of the shell between its open and closed states also makes it easy for users to know whether the microphone is capable of capturing audio or not.

8.2 Trust Issues Remain without Separation of Interests

While *Candid Mic* increases trust in a way that was accessible even to participants who self-reported as not being knowledgeable about electronics, it did not completely eliminate distrust. Some participants remained skeptical — they thought it might still be possible that the hinge disconnection was just a ruse and that the microphone could remain active even when the cover was closed. In other words, a visible physical disconnection alone may not fully provide perceptible assurance to all users. There is also a need to educate users as to the mechanics of microphone operation such that designs that provide perceptible assurance can engender trust. Future work can explore educational interventions to that end.

Trust is a social phenomenon unlikely to be solved with only a technical solution when distrust runs deep. As discussed by prior work [40], people's trust in whether their microphone is eavesdropping could be correlated with how much they trust the microphone manufacturer. Indeed, several of our participants mentioned that if they don't trust the microphone manufacturer, they will remain distrustful even after physical inspection. Thus, in addition to perceptible assurance and intentional powering, we argue that there is a need for a separation of interests: third-party microphones that interface with otherwise sensorless smart speakers.

8.3 Towards Practical Deployability

We engineered a proof-of-concept version of *Candid Mic* that works, but is sensitive to environmental conditions. For something like *Candid Mic* to be practically deployable, a number of engineering problems must be solved that make it more resilient, reliable, and durable. Self-sustainable technologies are increasingly important and in demand, so we believe that there will soon come an inflection point where there will be enough institutional engineering knowledge about how to engineer these devices in a manner that is resilient, reliant, and durable. With studies such as ours, when that day comes, we

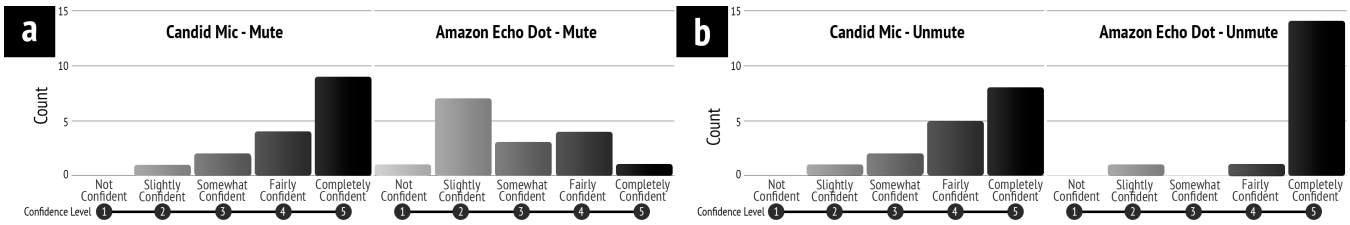

Figure 5: Quantitative data results about the confidence level of the following statements: (a) "When the microphone is muted, the voice assistant's microphone has no ability to hear what I am saying."; and (b) "When the microphone is unmuted, the voice assistant's microphone has the ability to hear what I am saying." Participants trusted that *Candid Mic* cannot record private conversations when muted significantly more than the Amazon Echo Dot. On the other hand, participants were generally confident that both *Candid Mic* and the Amazon Echo Dot can record private conversations when unmuted.

will be ready to engineer these devices in a way that is not only functional but accepted by users.

8.4 Study Limitations and Future Work

One limitation in our summative evaluation was that we did not capture participants' impressions of *Candid Mic* prior to watching an explainer video — though participants were able to use, manipulate, and ask questions about it prior to watching the video. Also, while explicitly showing the physical disconnection between the microphone and power source for *Candid Mic*, participants were not directly informed during the study that the Amazon Echo devices have internal hardware kill switches, which could have created a knowledge gap between the two conditions. In turn, this gap could have resulted in lower trust ratings for the Amazon Echo Dot.

In addition, our summative evaluation was an in-lab study with 16 participants. While this participant size is parred for the course in human-centered summative evaluations [8], our participant pool over-represents those with a background in computer science. Thus, we look forward to running an "in-the-wild" study with a larger, more census-representative sample of users to evaluate how *Candid Mic* affects user trust over the long term. However, for this long-term study, there are several challenges to address. First, our current prototype requires a backscatter infrastructure that consists of an RF transmitter and receiver, which are uncommon at present to have in everyday environments. Second, powering the prototype is susceptible to lighting as our energy harvesting source is limited to ambient light by photodiodes. To that end, our prototype is unlikely to work properly in a dark environment. While the long-term field study is not immediately practical, this work positions us and the community well to undertake such an effort when the technology is more practical.

8.5 Design Opportunities for 'Intentionally-powered' Sensing Devices

Candid Mic had a number of features that users liked from the usability perspective. Chief among these was portability — users appreciated not only having access to their smart speaker in one location. However, there were several usability challenges as well. One challenge is that our clamshell design — which prioritizes perceptible assurance — is not hands-free or seamless. Seamlessness is a fraught design value that often conflicts with other design values such as agency and privacy [9]. One fascinating opportunity for future work is to explore the relationship between the usability benefits of seamlessness and the privacy benefits of intentional powering in the realm of smart speaker microphone design. We suspect that different users will fall at different points of the spectrum between their desire for seamlessness versus intentional powering, and that there may be room for myriad design solutions to that end. It would also be pertinent to explore ways to allow for hands-free interactions that still enable intentional powering.

More generally beyond smart speaker microphones, *Candid Mic* is an example of an "intentionally-powered" sensor that aligns sensor use with clear knowledge of its use, and sensor disuse with trust in its disuse. As surveillance and distrust in institutions as stewards of personal data rise, there is a vast opportunity for exploring the design space of "intentionally-powered" sensing devices. Future work could, for example, explore how to create intentionally-powered RFID/NFC sensors, smart doorbells, and a vast array of other sensing technologies that allow users to unlock the utility of smart environments without the need of giving up all their privacy in so doing.

9 Conclusion

Many users do not trust that existing smart speakers are muted even if they are set to mute. Indeed, through formative interviews with 10 participants who have privacy concerns with smart speakers, we found that current smart speaker designs fail to provide users with perceptible assurance that the mi-

crophone is not listening in line with user intentions and expectations. We designed and implemented an "intentionally-powered" smart speaker microphone that allows end-users to visibly verify that the microphone is physically disconnected from its self-powered power source when not in explicit use, providing perceptible assurance. We ran a within-subjects experiment with 16 participants who have privacy concerns about covert recordings by smart speakers, comparing our new design against an Amazon Echo Dot in terms of the trust. We found that users had more trust in our re-designed microphone due to its being intentionally powered and providing perceptible assurance. Our work opens up the design space for a new class of intentionally-powered sensors that allow users to unlock the utility of smart environments without needing to compromise their privacy preferences.

Acknowledgments

This work was generously supported, in part, by Cisco and the National Science Foundation through grant SaTC 2029519. We are grateful to Anna Osterholm and John Reynolds for helping us fabricate the display. We thank the GVU Prototyping Lab for providing the fabrication facility. We would also like to thank members of the SPUD lab who provided valuable feedback throughout our research. Finally, we thank our anonymous reviewers for their feedback.

References

[1] Imtiaz Ahmad, Taslima Akter, Zachary Buher, Rosta Farzan, Apu Kapadia, and Adam J Lee. Tangible privacy for smart voice assistants: Bystanders' perceptions of physical device controls. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2):1–31, 2022.

[2] Imtiaz Ahmad, Rosta Farzan, Apu Kapadia, and Adam J Lee. Tangible privacy: Towards user-centric sensor designs for bystander privacy. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2):1–28, 2020.

[3] Apple. About the orange and green indicators in your iphone status bar, Dec 2022. https://support.apple.com/en-us/HT211876 (Accessed on 01/31/2023).

[4] Nivedita Arora, Ali Mirzazadeh, Injoo Moon, Charles Ramey, Yuhui Zhao, Daniela C Rodriguez, Gregory D Abowd, and Thad Starner. Mars: Nano-power battery-free wireless interfaces for touch, swipe and speech input. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, pages 1305–1325, 2021.

[5] Nivedita Arora, Steven L Zhang, Fereshteh Shahmiri, Diego Osorio, Yi-Cheng Wang, Mohit Gupta, Zhengjun Wang, Thad Starner, Zhong Lin Wang, and Gregory D Abowd. Saturn: A thin and flexible self-powered microphone leveraging triboelectric nanogenerator. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(2):1–28, 2018.

[6] Virginia Braun and Victoria Clarke. *Thematic analysis.* American Psychological Association, 2012.

[7] Matthew Brocker and Stephen Checkoway. {iSeeYou}: Disabling the {MacBook} webcam indicator {LED}. In *23rd USENIX Security Symposium (USENIX Security 14)*, pages 337–352, 2014.

[8] Kelly Caine. Local standards for sample size at chi. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 981–992, 2016.

[9] Matthew Chalmers, Ian MacColl, and Marek Bell. Seamful design: Showing the seams in wearable computing. In *2003 IEE Eurowearable*, pages 11–16. IET, 2003.

[10] Varun Chandrasekaran, Suman Banerjee, Bilge Mutlu, and Kassem Fawaz. {PowerCut} and obfuscator: An exploration of the design space for {Privacy-Preserving} interventions for smart speakers. In *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*, pages 535–552, 2021.

[11] Batyr Charyyev and Mehmet Hadi Gunes. Misactivation detection and user identification in smart home speakers using traffic flow features. In *Proceedings of the 14th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, pages 135–146, 2021.

[12] Yuxin Chen, Huiying Li, Shan-Yuan Teng, Steven Nagels, Zhijing Li, Pedro Lopes, Ben Y Zhao, and Haitao Zheng. Wearable microphone jamming. In *Proceedings of the 2020 chi conference on human factors in computing systems*, pages 1–12, 2020.

[13] Long Cheng, Fang Liu, and Danfeng Yao. Enterprise data breach: causes, challenges, prevention, and future directions. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(5):e1211, 2017.

[14] Ming Ki Chong and Hans Gellersen. Usability classification for spontaneous device association. *Personal and Ubiquitous Computing*, 16(1):77–89, 2012.

[15] Juliet Corbin and Anselm Strauss. *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage publications, 2014.

[16] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. Wizard of oz studies: why and how. In *Proceedings of the 1st international conference on Intelligent user interfaces*, pages 193–200, 1993.

[17] Youngwook Do, Jung Wook Park, Yuxi Wu, Avinandan Basu, Dingtian Zhang, Gregory D Abowd, and Sauvik Das. Smart webcam cover: Exploring the design of an intelligent webcam cover to improve usability and trust. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(4):1–21, 2021.

[18] Daniel J Dubois, Roman Kolcun, Anna Maria Mandalari, Muhammad Talha Paracha, David Choffnes, and Hamed Haddadi. When speakers are all ears: Characterizing misactivations of iot smart speakers. *Proceedings on Privacy Enhancing Technologies*, 2020(4):255–276, 2020.

[19] Serge Egelman, David Molnar, Nicolas Christin, Alessandro Acquisti, Cormac Herley, and Shriram Krishnamurthi. Please continue to hold. In *Ninth Workshop on the Economics of Information Security*, 2010.

[20] ACB English. Zeff bezos - the vanity fair new establishment summit with amazon ceo and walter isaacson., Sep 2021. https://www.youtube.com/watch?v=5UGwFTdAk3I (Accessed on 05/23/2023).

[21] Elin L. Howard, Anna M. Österholm, D. Eric Shen, L. Prerana Panchumarti, Carlos Pinheiro, and John R. Reynolds. Cost-effective, flexible, and colorful dynamic displays: Removing underlying conducting layers from polymer-based electrochromic devices. *ACS Applied Materials & Interfaces*, 13(14):16732–16743, 2021. PMID: 33788540.

[22] Yue Huang, Borke Obada-Obieh, and Konstantin Beznosov. Amazon vs. my brother: How users of shared smart speakers perceive and cope with privacy risks. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–13, 2020.

[23] Haojian Jin, Boyuan Guo, Rituparna Roychoudhury, Yaxing Yao, Swarun Kumar, Yuvraj Agarwal, and Jason I Hong. Exploring the needs of users for supporting privacy-protective behaviors in smart homes. In *CHI Conference on Human Factors in Computing Systems*, pages 1–19, 2022.

[24] Bjørn Karmann and Tore Knudsen. Project alias, 2018. https://bjoernkarmann.dk/project_alias (Accessed on 01/31/2023).

[25] Deepak Kumar, Riccardo Paccagnella, Paul Murley, Eric Hennenfent, Joshua Mason, Adam Bates, and Michael Bailey. Skill squatting attacks on amazon alexa. In *27th {USENIX} Security Symposium ({USENIX} Security 18)*, pages 33–47, 2018.

[26] Marc Langheinrich. Privacy by design—principles of privacy-aware ubiquitous systems. In *International conference on ubiquitous computing*, pages 273–291. Springer, 2001.

[27] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. Alexa, are you listening? privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1–31, 2018.

[28] Kathleen M MacQueen, Eleanor McLellan, Kelly Kay, and Bobby Milstein. Codebook development for team-based qualitative analysis. *Cam Journal*, 10(2):31–36, 1998.

[29] Nathan Malkin, Joe Deatrick, Allen Tong, Primal Wijesekera, Serge Egelman, and David Wagner. Privacy attitudes of smart speaker users. *Proceedings on Privacy Enhancing Technologies*, 2019(4):250–271, 2019.

[30] Abraham Mhaidli, Manikandan Kandadai Venkatesh, Yixin Zou, and Florian Schaub. Listen only when spoken to: Interpersonal communication cues as smart speaker privacy controls. *Proceedings on Privacy Enhancing Technologies*, 2020(2):251–270, 2020.

[31] Richard Mitev, Anna Pazii, Markus Miettinen, William Enck, and Ahmad-Reza Sadeghi. Leakypick: Iot audio spy detector. In *Annual Computer Security Applications Conference*, pages 694–705, 2020.

[32] Roger J Mortimer, Aubrey L Dyer, and John R Reynolds. Electrochromic organic and polymeric materials for display applications. *Displays*, 27(1):2–18, 2006.

[33] Jakob Nielsen. Finding usability problems through heuristic evaluation. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 373–380, 1992.

[34] Don Norman. *The design of everyday things: Revised and expanded edition*. Basic books, 2013.

[35] NPR and Edison Research. The smart audio report, June 2022. https://www.nationalpublicmedia.com/uploads/2020/04/The-Smart-Audio-Report_Spring-2020.pdf (Accessed on 01/31/2023).

[36] Rebecca S Portnoff, Linda N Lee, Serge Egelman, Pratyush Mishra, Derek Leung, and David Wagner. Somebody's watching me? assessing the effectiveness of webcam indicator lights. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1649–1658, 2015.

[37] Jason Pridmore and Anouk Mols. Personal choices and situated data: Privacy negotiations and the acceptance of household intelligent personal assistants. *Big Data & Society*, 7(1):2053951719891748, 2020.

[38] Jason Pridmore, Michael Zimmer, Jessica Vitak, Anouk Mols, Daniel Trottier, Priya C Kumar, and Yuting Liao.

Intelligent personal assistants and the intercultural negotiations of dataveillance in platformed households. *Surveillance & Society*, 2019.

[39] Nirupam Roy, Haitham Hassanieh, and Romit Roy Choudhury. Backdoor: Making microphones hear inaudible sounds. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, pages 2–14, 2017.

[40] Christine Satchell and Paul Dourish. Beyond the user: use and non-use in hci. In *Proceedings of the 21st annual conference of the Australian computer-human interaction special interest group: Design: Open 24/7*, pages 9–16, 2009.

[41] Lea Schönherr, Maximilian Golla, Thorsten Eisenhofer, Jan Wiele, Dorothea Kolossa, and Thorsten Holz. Unacceptable, where is my privacy? exploring accidental triggers of smart speakers. *arXiv preprint arXiv:2008.00508*, 2020.

[42] Alex Sciuto, Arnita Saini, Jodi Forlizzi, and Jason I Hong. " hey alexa, what's up?" a mixed-methods studies of in-home conversational agent usage. In *Proceedings of the 2018 designing interactive systems conference*, pages 857–868, 2018.

[43] Yunpeng Song, Yun Huang, Zhongmin Cai, and Jason I Hong. I'm all eyes and ears: Exploring effective locators for privacy awareness in iot scenarios. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2020.

[44] Statista. Market share of global smart speaker shipments from 3rd quarter 2016 to 2nd quarter 2021, by vendor, 2022. (Accessed on 11/15/2022).

[45] Statista. Reasons for not owning a smart speaker according to online users in the united states as of february 2020, Mar 2022. https://www.statista.com/statistics/1059646/us-reasons-for-not-owning-a-smart-speaker/ (Accessed on 09/09/2022).

[46] Ke Sun, Chen Chen, and Xinyu Zhang. " alexa, stop spying on me!" speech privacy protection against voice assistants. In *Proceedings of the 18th conference on embedded networked sensor systems*, pages 298–311, 2020.

[47] Alistair Sutcliffe. Trust: From cognition to conceptual models and design. In *Advanced Information Systems Engineering: 18th International Conference, CAiSE 2006, Luxembourg, Luxembourg, June 5-9, 2006. Proceedings 18*, pages 3–17. Springer, 2006.

[48] Madiha Tabassum, Tomasz Kosinski, and Heather Richter Lipford. I don't own the data": End user perceptions of smart home device data practices and risks. In *SOUPS@ USENIX Security Symposium*, 2019.

[49] Yu-Chih Tung and Kang G Shin. Exploiting sound masking for audio privacy in smartphones. In *Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security*, pages 257–268, 2019.

[50] Tavish Vaidya, Yuankai Zhang, Micah Sherr, and Clay Shields. Cocaine noodles: exploiting the gap between human and machine speech recognition. In *9th {USENIX} Workshop on Offensive Technologies ({WOOT} 15)*, 2015.

[51] Naomi 'SexyCyborg' Wu. Diy wearable microphone jammer!, Mar 2021. https://youtu.be/H1rozZ7ebxQ (Accessed on 01/31/2023).

[52] Yucheng Yang, Jack West, George K Thiruvathukal, Neil Klingensmith, and Kassem Fawaz. Are you really muted?: A privacy analysis of mute buttons in video conferencing apps. *arXiv preprint arXiv:2204.06128*, 2022.

[53] Yaxing Yao, Justin Reed Basdeo, Smirity Kaushik, and Yang Wang. Defending my castle: A co-design study of privacy mechanisms for smart homes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2019.

[54] Eric Zeng, Shrirang Mare, and Franziska Roesner. End user security and privacy concerns with smart homes. In *Symposium on Usable Privacy and Security (SOUPS)*, volume 220, 2017.

Appendix

A Study Material

A.1 Formative Study

A.1.1 Pre-screener

We used Prolific's pre-screening features first by geographic location. Then, we used Quatrics questions that ask about the main reason for smart speaker non-adoption as follows:

- Do you currently use a smart speaker device (e.g., Amazon Echo, Apple Siri, Google Nest, Facebook Portal, etc.)?
- Have you owned or used a smart speaker before?
- What is your primary reason for currently not using a smart speaker?

  - Lack of utility
  - Privacy concerns
  - Expensive product price
  - Others

A.1.2 Interview Questionnaire

The interview questions consist of two parts: (1) contextualizing questions; and (2) interaction design questions. The contextualizing question is for participants to warm up and feel comfortable talking about their experience. Then, we moved on to the main part of the study by asking the questions about interaction design of our mock-up prototypes.

Contextualizing Questions

- Have you ever owned a smart speaker previously that you stopped using?
  - (If they answer yes) What did you typically use it for? Can you give me a specific example?
  - Why did you stop using it?
    - (If they mention privacy) You mentioned that privacy was one of your reasons, can you tell me more about that?
    - (If they mention distrust in large companies) Why do you not trust smart speakers/companies?
    - (If they do not bring up privacy) In the first survey, you selected privacy as a reason for not adopting a smart speaker, what concerns do you have?
  - (If they answer no) From the screener, you indicated that you do not currently own a smart speaker. Had you used or interacted with a smart speaker before you made your decision not to get one? Have you ever considered buying one? Can you tell us more about why you decided to not buy one?
    - (If they mention privacy) You mentioned that privacy was one of your reasons, can you tell me more about that?
    - (If they mention distrust in large companies) Why do you not trust smart speakers/companies?
    - (If they do not bring up privacy) In the first survey, you selected privacy as a reason for not adopting a smart speaker, what concerns do you have? Why do you have privacy concerns?

Interaction Design Questions We randomly ordered three mock-up prototypes' interactions. Then, we asked participants to watch the video of the first prototype interaction and asked a set of questions. Afterward, we applied the same protocol to the second and third prototypes separately.

- What do you like/dislike about this design?
  - (If someone asks technical questions) are those important design considerations to you? Why?
- (Comparison with the Qualtrics questionnaire) This microphone is turned on only during this interaction. How confident are you with the following statement: the microphone is not listening to you when you don't want the microphone to listen?
  - Likert Scale
    - Not confident at all
    - Slightly confident
    - Somewhat confident
    - Fairly confident
    - Completely confident
  - (For the first prototype) Why did you choose this score? (For the second and third prototypes) Why did you choose the score differently (or same) from the others?
- (Suggestions) Do you have any other thoughts or ideas to improve interactions?

A.2 Summative Study

A.2.1 Pre-screener

We pre-screen participants via email. We confirmed our eligibility criteria by asking the following categories:

- Participants in this study must: (1) be at least 18 years old; (2) be fluent in English; (3) be able to participate in our study in person; and (4) either currently NOT use any smart speaker devices because of privacy concerns, or be a smart speaker user who has privacy concerns against a smart speaker.

A.2.2 Interview Questionnaire

We structured the interview questions split into two parts same as the formative study interview questions: (1) contextualizing questions; and (2) interaction design questions. After asking the contextualizing questions, we showed participants the video that illustrates how each condition's device works and let them directly play with them. Then, we asked the interaction questions.

Contextualizing Questions

- Have you used or owned a smart speaker before? Why?
- Could you tell me why you have privacy concerns?

Interaction Questions

- (General reactions) What do you like/dislike about it?
- (Comparison with the Qualtrics questionnaire) How confident are you with the following statements? "When the microphone is muted, the voice assistant's microphone has no ability to hear what I am saying."
  - Likert Scale
    - Not confident at all
    - Slightly confident
    - Somewhat confident
    - Fairly confident
    - Completely confident
  - (For the first prototype) Why did you choose this score? (For the second and third prototypes) Why did you rate this differently (or same) from the others?
- (Comparison with the Qualtrics questionnaire) How confident are you with the following statements? "When the microphone is unmuted, the voice assistant's microphone has the ability to hear what I am saying."
  - Likert Scale
    - Not confident at all
    - Slightly confident
    - Somewhat confident
    - Fairly confident
    - Completely confident
  - (For the first prototype) Why did you choose this score? (For the second prototype) Why did you rate this differently (or same) from the others?
  - (Preference questions) Which would you prefer to use between a new smart speaker and Alexa? Why?

B *Candid Mic* Implementation Detail

B.1 Low-power Electrochromic Display (ECD)

The ECD device design consists of two electrodes – the working and counter electrodes, and an electrolyte gel sandwiched in between. The working has a coating of pink-colored electrochromic ink called ECP-magenta on a conductive PET sheet, and the counter electrode has a layer of transparent nano-indium tin oxide (ITO) coated on the same.

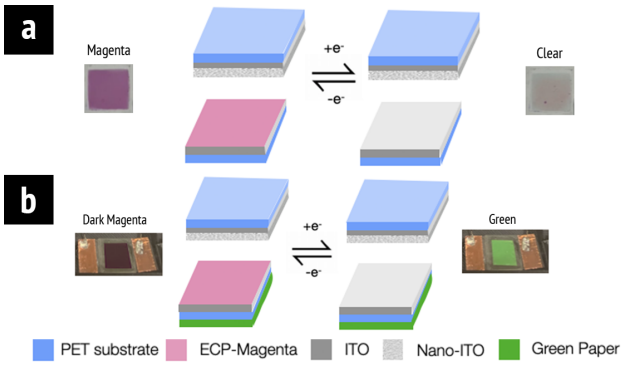

Figure 6: This figure illustrates how the ECD is fabricated. (a) ECP-magenta switches its color between magenta and clear. (b) After placing a green sheet of paper, we made it seem to switch its color from dark magenta and green.

Electrically, this device can be considered a supercapacitor. ECD is magenta color at default 0 voltage. The device changes color from magenta to colorless with the application of appropriate electrode potential due to redox chemical reaction that occurs between the two electrodes [32] (See Figure 6 (a)). When the two electrodes are short-circuited, the ECD discharges and returns to its default magenta color state.

Typically, the switching time of a traditional display refers to the amount of time it takes to reach 95% of its full intensity switch. In the case of ECP-Magenta, we use the time for ECDs to complete 95% of its full contrast switch, the switching time for oxidation is 0.35 seconds (from colorless to magenta), and 0.45 seconds for reduction (from magenta to colorless) when the applied voltage is $\pm 0.45V$.

To enhance the gamut of colors that the ECD device can express, a green-colored paper is added below the device. We chose green as a color that represents an on-active state. In default state with 0V, this device (See Figure 6 (b)) is dark magenta that represents the off-inactive state of wireless communication. When Voltage is applied it turns green.

B.2 Circuit Design

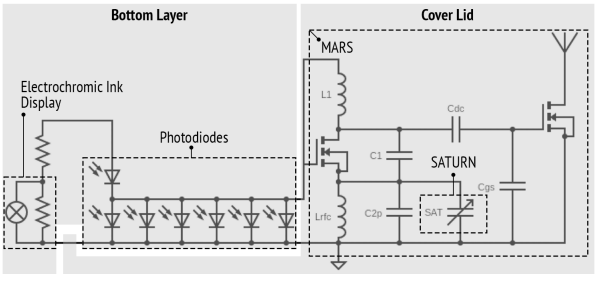

Figure 7: This figure shows the schematic design and the wire arrangement on the cover and bottom layers of *Candid Mic*.