PowerCut and Obfuscator: An Exploration of the Design Space for Privacy-Preserving Interventions for Smart Speakers

Varun Chandrasekaran, Suman Banerjee, Bilge Mutlu, Kassem Fawaz
University of Wisconsin-Madison

Abstract

The pervasive use of smart speakers has raised numerous privacy concerns. While work to date provides an understanding of user perceptions of these threats, limited research focuses on how we can mitigate these concerns, either through redesigning the smart speaker or through dedicated privacy-preserving interventions. In this paper, we present the design and prototyping of two privacy-preserving interventions: 'Obfuscator' targeted at disabling recording at the microphones, and 'PowerCut' targeted at disabling power to the smart speaker. We present our findings from a technology probe study involving 24 households that interacted with our prototypes; the primary objective was to gain a better understanding of the design space for technological interventions that might address these concerns. Our data and findings reveal complex trade-offs among utility, privacy, and usability and stresses the importance of multi-functionality, aesthetics, ease-of-use, and form factor. We discuss the implications of our findings for the development of subsequent interventions and the future design of smart speakers.

1 Introduction

Smart speakers, or network-connected speakers with integrated virtual assistants, are becoming increasingly pervasive in households. In 2020, nearly 90 million US adults used a smart speaker [44]. Smart speakers offer their users a convenient way to access information, set alarms, play games, or set to-do lists. Smart speakers also integrate with other devices to realize smart home applications. However, this convenience comes at a potential privacy cost; these devices operate in an always-on mode at earshot of nearby conversations.

Figure 1: The Obfuscator design probe next to a Google Home Mini Device. Obfuscator uses ultrasound jamming to prevent the smart speaker from listening to the user's conversations and is designed to appear as a tabletop "trinket" to blend into the user's home environment.

Smart speakers already provision built-in privacy controls; they are supposed to process audio inputs locally until they detect a wake word, and they pack a button that mutes their internal microphone. Unfortunately, both provisions are not very effective at protecting the user's privacy. Recent incidents raise concerns about *passive* privacy threats [1, 22, 26, 27]. Smart speakers can be mistakenly triggered without the presence of a wake word [11, 19, 25], causing it to record speech not intended as commands. Further, security researchers have documented *active* vulnerabilities that indicate the potential for malicious exploitation of smart speakers [10, 16, 18, 34, 48]. Further, the effectiveness of the mute button to address these problems is in doubt [26]. Recent studies, including the one in this work, indicate that users find this button inconvenient to utilize and not trustworthy in some cases [42]. While different technical interventions have been proposed recently [8, 42], the design space for such interventions remains under-explored. This paper contributes to an improved understanding of the design elements and understanding user

USENIX Symposium on Usable Privacy and Security (SOUPS) 2021.
August 8–10, 2021, Virtual Conference.

experience with these interventions.

In our work, we aim to understand better the user perceptions around the potential technological solutions to the privacy issues involving smart speakers through a technology probe-based approach [21]. The objective of our study is not to validate particular design choices but to understand user perceptions of such interventions better and extract design requirements for them. We utilize the smart speaker's *built-in* mute button as a baseline, to understand user perceptions of how device manufacturers provide privacy control. We utilized two technology probes to represent *bolt-on* privacy-preserving interventions: (a) PowerCut, a smart plug that allows the user to engage/disengage the power supply to a smart speaker remotely, and (b) Obfuscator (Figure 1), which uses ultrasound to deafen the smart speaker's microphone, preventing the smart speaker from listening to nearby conversations. The probes intercept two key resources required for successful smart speaker functionality: *power* (for basic operation) and *microphone inputs* (for voice-based interaction).

To promote user reflection on our privacy-preserving interventions, we conducted in-home demonstrations of our technology probes through in-depth interviews at 24 households. Our interviews took place over two phases between July 2018 and August 2019 , providing us with insight into how such perceptions and attitudes might change over time. Our interviews involved users with diverse demographics, including *casual* (or recreational) users and *power* (or proficient) users, enabling us to distinguish perceptions and design requirements for different user groups. Our findings highlight a complex trade-off between privacy, utility, and usability: the interventions (a) should be plug-and-playable *i.e.,* require minimal setup and upkeep, (b) have a small physical footprint and fit within its environment, (c) offer additional features beyond privacy preservation, (d) does not affect the interaction model with the smart speaker, and (e) must survive the test of time *i.e.,* it should be compatible with existing and future iterations of smart speakers. Through this work, we present the design of our technology probes and our in-home study, and discuss our findings. We conclude with a discussion of their implications for the smart speakers as well as other privacy-sensitive technologies.

2 Background

Our study considers smart speakers deployed in home environments, focusing on (a) Google Home Mini and (b) Amazon Echo Dot as described in Table 1. Users interact with these devices to achieve a multitude of tasks, such as information access, interaction with other smart devices, setting alarms/timers, and voice calls. A typical interaction with a smart speaker starts with the user speaking a wake word, such as "*...Alexa*" or "*...Google.*" Upon recognition of the wake word, the device indicates its readiness to receive command through a visual cue. Then, the device sends the speech segment to the cloud, which verifies the wake word and processes the accompanying command [40]. Verification is necessary since on-device models are typically less accurate to minimize their compute footprint and latency of predictions [32, 41]. As such, the smart speaker has to be always on, continuously listening for a user to speak the wake word. Ideally, the device should only record, and communicate to its cloud, the commands that were triggered by a wake word. In many circumstances, however, the device's operation might not match its expected behavior. This results in the two privacy threats described below. Note, these threats also provide context about scenarios where we envision privacy-preserving interventions to be used.

Feature	Home Mini	Echo Dot
Manufacturer	Google	Amazon
Height × Diameter	4.3 × 9.9 cm	3.3 × 7.6 cm
Wake words	"... Google"	"... Alexa"
Visual Cue	Dots on the surface	LED band
Privacy controls	Mute button	On/Off switch

Table 1: Salient features of smart speakers in 2019.

Passive Threats: The first threat occurs due to innocuous and inadvertent recording *i.e.,* when the smart speaker misunderstands ongoing conversations to contain the wake word. Recent analysis [11] reported that everyday phrases, such as those from TV shows, can accidentally activate a smart speaker, resulting in 10 seconds of speech being sent to the cloud. There have been several incidents where these devices have exported user conversation, including those not preceded with a wake word. While one organization claims this is a one-off act [19], another blames erroneous code [10, 15]. There have also been reported instances where several organizations hired human contractors to listen and tag different recordings from these devices, which include commands and non-commands [16]; this is a severe deviation from perceived device operation. Collectively, we refer to these violations as *passive* privacy threats.

Active Threats: The second occurs due to compromise of the actual device or its operation. A malicious entity can compromise the software running on the connected smart speaker to turn it into a listening device. Such an entity can also change the operation of the device through developing applications that record the user's conversations [25, 34, 49] or inject stealthy commands to wake up the device without the user's awareness [6, 38, 48]. Since these devices are connected to the internet, such alterations are capable of extracting various forms of sensitive information. We refer to such threats as *active* privacy threats.

3 Methodology

We envision privacy-preserving interventions to address potential passive and active threats, especially in scenarios that users perceive as sensitive. Such scenarios can include users receiving visitors or having sensitive conversations. Concretely, there exist two strategies to safeguard users' privacy in sensitive scenarios: (a) redesigning the smart speaker to provide provable privacy guarantees, or (b) designing interventions that co-exist with the smart speaker. The former is a challenging proposition as most of the software and hardware required for successful smart speaker functioning is proprietary. Additionally, it would involve trusting the device provider (a theme that will revisit later) to provide proof that the user's privacy was not violated.

To this end, we explore the design of *bolt-on, hardware-based* interventions. These interventions are less abstract than software-based ones; they allow the users to physically and directly interact with them. For thoroughness, we compare and contrast our findings with the usage of a *built-in* feature found in smart speakers— the mute button. The results of our research inform the design of smart speakers with improved privacy properties and privacy-preserving interventions in physical spaces. Note that our analysis is restricted to smart speakers and not smartphones (which are also susceptible to the threats discussed earlier). In particular, smart speakers are *easier to protect* as they are less *mobile* than smart-phones.

What is a tech probe? We follow a *technology probe*-based design approach, which allows us to identify design guidelines that capture the users' mental models. We aim to understand how the users of smart speakers react to different privacy-enhancing technologies using proof-of-concept prototypes (or probes). In a technology probe, the researcher develops an interface that packages the core functionality of the privacy intervention. The researcher keeps the interface as simple as possible to avoid making design decisions [5, 33]. When an individual interacts with this basic interface, the researcher *probes* the individual to reveal a specific phenomenon that is otherwise hidden [21].

In our case, we probe and interview the users to elicit their immediate reactions and reflections about what design elements are missing and need to be introduced. We follow with qualitative analysis to reveal the design guidelines for a privacy intervention in the smart speaker environment. In follow-up work, we are planning to realize the privacy intervention and set up a diary study to understand longer-term use. This will allow us to concretely measure any issues users have with the actual intervention that was conceptualized for deployment. A note on the nomenclature: in this work, we design technology probes (or probes for short) to elicit insight about the final intervention (which we do not design), for which we make recommendations.

3.1 Iterative Design Process

In designing our technology probes, we followed an iterative design process. We first explored the broad space of solutions (presented in Table 2), their efficacy against an adaptive adversary, and discussed the advantages and disadvantages of each approach. Recall that our objectives are to design an easy-to-use intervention with intuitive yet provable privacy guarantees. It is clear that modifying device hardware and controlling network flow does not provide the desired privacy protection – the encrypted nature of network traffic makes it difficult to tag and discard packets (with information) that are not to be shared, while inadvertent smart speaker activation will persist. One could change the wake word to reduce the frequency of spurious activation/recording. However, this phenomenon is not well understood for it to be a definitive fix, and a harder-to-pronounce wake word has usability problems.

Possible Solutions	***Active* Threats**	***Passive* Threats**
Network interception	✗	✗
Hardware modifications	✗	✗
Change the wake word	✗	✓
Discard smart speaker	✓	✓

Table 2: Space of possible solutions and their effectiveness against malicious programming (or *active* privacy threats) and inadvertent recording (or *passive* privacy threats).

Observe that while some of our possible solutions are intuitive to the average user, others (such as network monitoring) are not. Based on preliminary discussion with several end-users, we converged on a set of dimensions that we found relevant to the final design of our probes. They are (a) the method of user-probe interaction *i.e.,* hands-free vs. physical, (b) the ease of deployment, and (c) the ease of understanding the privacy properties the probe provides. We stress that these dimensions are not exhaustive and merely serve as a starting point for our design.

We construct two probes guided by these suggestions. Again, we stress that we do not seek to evaluate the efficacy of these probes in preserving privacy. We do not attempt to understand how people use these probes as well. Doing so requires running a diary study with the probe deployed in users' homes. We describe the probes used in our study, including those we conceptualized, below. We also briefly state our analysis of the trade-offs ensuing from each probe.

1. Mute: The "mute" feature represents a *built-in* privacy control (Figure 2a). It is available as a push button on the top panel of some of the Amazon Echo Dots and as a sliding button on the side of some of the Google Home Minis. The device manufacturers state that the microphone is deactivated when the mute button is turned on (*c.f.* Figure 2a). Naturally, activating the mute button stops the smart speaker from re-

sponding to the user's voice commands. Upon activation, the Echo Dot's ring color changes to red, and the four lights atop the Google Home Mini turn red.
Trade-offs: While inbuilt, the mute feature requires the user to physically interact with the device to engage the control. It also requires the user to place trust in the manufacturer's implementation of the feature.

2. PowerCut: While the mute button focuses on disengaging the microphone inputs, we conceptualize another probe to disengage the electricity supply. A naive way of achieving our goal is to either disconnect the smart speaker's cord from the outlet or disconnect the cord connected to the smart speaker. However, both options involve physical interaction with the device. Thus, we use a remote-controlled outlet[1] (Figure 2b). The user deploys PowerCut by connecting the smart speaker to the outlet through the smart plug (as seen in Figure 2b).
Trade-offs: We use a commercial smart-plug because we believe that users will be familiar with such products, minimizing their time for acclimatization. Additionally, we speculate that users will trust the functionality of such widely-used products, with no negative publicity. PowerCut is conspicuous and rugged; we believe that its form factor makes it easier to understand and use. The user can engage/disengage PowerCut through a remote control (with a range of operation of 100 feet) without the need to physically interact with the device. Additionally, the smart plug we chose provides a visual cue — an LED glows *red* when powered on to indicate that the smart speaker is active. Clearly, PowerCut offers immediate privacy guarantees. This comes at a cost; the users have to wait for a lengthy boot time whenever they wish to reuse the smart speaker. Additionally, the form factor of PowerCut makes it difficult to use in some environments (with concealed/narrow outlets).

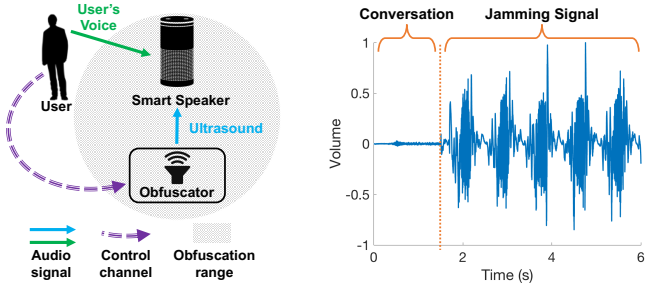
3. Obfuscator: This probe targets the microphone of the smart speaker (Figure 2c). Obfuscator generates *inaudible* ultrasound to deafen the microphone of the smart speaker when the user needs privacy protection (Figure 3a). Using a remote control, users are able to engage/disengage the probe without having to physically interact with it. When disengaging the jamming, the user can *immediately* interact with the smart speaker. Due to non-linearities in off-the-shelf microphones' power and diaphragm [13,37,38,48], Obfuscator creates high-power, human-inaudible noise at these microphones but does not affect its operation. Figure 3b shows the captured signals from a commodity microphone before and after Obfuscator is engaged. Before jamming is invoked, the microphone records a conversation, which is audible at playback. After engaging Obfuscator, the ultrasound jamming signal is recorded at the microphone and completely overwhelms the conversation's signal. The circuitry of Obfuscator includes a remote-controlled DC power supply, an ultrasound generator, and a horn speaker that emits the ultrasound signal.

The design of Obfuscator utilizes a jamming signal with randomized tones at the ultrasound frequency range, which manifest as randomized tones at the audible range. Theoretically, a determined smart speaker manufacturer can attempt to filter these tones at the expense of a degraded speech signal; such degradation might result in a deteriorated performance of wake word detection, which hinders the utility of the smart speaker. Our experiments show that the jamming from Obfuscator is effective at blocking the wake word detection.

[1] Beastron Remote Controlled Outlet

(a) Mute (b) PowerCut (c) Obfuscator

Figure 2: The three employed privacy probes.



(a) High-level overview (b) Microphone received signal

Figure 3: The system design of the Obfuscator probe.

3.1.1 Design Evolution

We explored different design options for the prototype that houses the circuitry. A challenge in prototyping Obfuscator was the footprint of the circuitry. Additionally, horn speakers are bulky, and reducing their size inhibits their efficacy. Our design process started with a search for a privacy metaphor, one that creates the perception of privacy control for the users. Our initial prototype was based on a "cage" metaphor. Here, the Obfuscator probe is housed in a cage-like structure with a door, and the smart speaker is placed within the cage. When the user closes the cage door, Obfuscator generates the ultrasound obfuscation signal to prevent the smart speaker from listening. The user has to manually open the door to disable obfuscation and communicate with the smart speaker. Closing the door "locks" the device in a cage, providing a user with a perception that the device is not active and their space is

private.

The first version of the Obfuscator probe followed the cage metaphor as a 3D printed cylinder (Figure 4a). The cylinder has two compartments; the lower chamber containing the circuit and the ultrasound speaker. The upper chamber has space for the smart speaker as well as the door. The first version has a height of 15.5 cm and a diameter of 12 cm. We refined this design into a lighter and less conspicuous 3D-printed cylinder (Figure 4b) with a height of 13 cm and a diameter of 11 cm. This was the second version.

Based on pilot studies with 2 participants, we found both versions to be neither user-friendly nor fitting with home decor. Participants explicitly indicated that this design was not something they would want in their homes. Further, we observed that individuals did not associate with the privacy metaphor. First, they did not favor the idea of physically interacting with the prototype as it takes away the convenience of using a hands-free device. Second, covering the smart speaker inside the cylinder deprives the users of the ability to observe the visual cue (refer Table 1). This is a shortcoming of placing the smart speaker within the probe. Finally, they thought that the actions of opening and closing the prototype door were conspicuous and would rattle others in the vicinity.

In the third version, we considered three aspects that the users were not fond of: physical interactions with the door, covering the smart speaker, and the aesthetics[2]. The third version of the prototype (Figure 4c) features a platform-like solution, which addresses those shortcomings. This version has a glass cylinder that houses the circuit and is covered by decorative sand; its height is 11 cm, and its diameter is 12.5 cm. The platform, where the smart speaker sits, is encased with synthetic leather. The user can engage/disengage the jamming signal via remote control, obviating the need for physical interaction. This version of the Obfuscator probe follows a different privacy metaphor: "virtual veil." By engaging the jamming signal, Obfuscator creates a virtual privacy dome around the smart speaker, preventing it from listening to the conversations. Our subsequent discussions and reflections about this version revealed that the open nature of the prototype might not enforce the privacy metaphor; users are less likely to perceive privacy control over the smart speaker. Additionally, this version remains co-located with the smart speaker, increasing its form factor. This is not ideal when the smart speaker is concealed.

[2]Aesthetics are subjective, and determining a good aesthetic for even a prototype is a challenging problem.

(a) V.1 (b) V.2 (c) V.3

Figure 4: The design evolution of the Obfuscator probe.

The design search process led to our final prototype of the Obfuscator probe, as shown in Figure 2c. We substantially reduced the form factor of the final version. The new prototype houses the same circuitry in a glass candle holder. The glass is filled with decorative sands and sealed with burgundy burlap. The user only needs to place the prototype next to the smart speaker. This prototype is built using commonly found household artifacts, enabling it to fit in with the existing decor. The final prototype (henceforth our probe) packages the core functionality of the privacy probe: a jamming device that enforces the privacy metaphor. We kept the prototype as simple and basic as possible to avoid making design decisions [5] that influence our findings. In our study, we use the prototype to elicit participants' reflections about what design elements are missing and need to be introduced.

3.2 In-home User Study

We recruited 24 families (including single individuals) within a 15-mile radius of the UW-Madison campus, utilizing the university mailing list, over two phases. Our first phase, in 2018, included 13 interviews, while the second phase (13 months later, in 2019) included 11 interviews. We use a 2-phased approach to obtain results from a wider variety of end-users; we wished to interview both unaware users (in phase 1) and those familiar with media reports of privacy violations induced by smart speaker, at the time of the interviews (in phase 2). We chose to perform shorter and focused interviews as opposed to longer studies (such as diary studies); the tech probe approach allowed us to capture our many goals related to capturing baseline privacy perceptions, introduce the privacy priming, and gain reflection upon interacting with the interventions.

The results reported in the paper are based on interviews with 30 participants ($P_1 - P_{30}$) from these 24 interviews[3]. Our data coding and analysis started immediately and took place simultaneously with data collection, enabling us to monitor the emergence of new codes and themes and determine saturation. We reached saturation by the 18^{th} interview and collected data from 6 more households to assess how perceptions evolve with time. Our approach exhibits several limitations, the most important of which is the sampling of a relatively (a) *ethnically* homogeneous and (b) educated population; the reported results are less likely to generalize to another population of users. We sought to recruit participants with different backgrounds in age, education, and technological proficiency. Our participant pool comprised 15 males and 15 females. The

[3]Some households had more than one participant.

youngest participant was 12 years old, while the oldest was 67 (with a mean age of 37.4 and a standard deviation of 13.9). The occupations of the participants ranged from students to faculty. The wide spectrum in age and profession enables us to gain feedback from a pool with varied technical knowledge and awareness and offers a breadth of experiences and backgrounds that are useful to analyze user interactions with the interventions.

We conducted all interviews at the participants' homes at a time of their convenience. Each interview lasted 90 minutes on average, and the participants were compensated for their time ($40 per study). The study protocol was approved by our Institutional Review Board. Each interview consisted of three stages, which we elaborate on below.

1. Environment Exploration: The interview began with the participants providing a brief tour of their home. Emphasis was placed on the rooms with smart speakers. Then, the interviewer and the participants convened in the room with the frequently used smart speaker so as to simulate a common usage scenario. After obtaining informed consent, the interviewer first asked the participants to interact with their smart speaker to ensure that it was operating as expected. This was followed by questioning participants about their knowledge/understanding of how smart speakers operate. Then, the interviewer asked more detailed questions about the smart speaker's role in the participant's life. The questions focused on frequency, duration, and the purpose of usage. Also, the questions covered the conversations and activities participants perform around their smart speakers. Then, the interviewer inquired about the participant's degree of trust in these devices (in terms of the potential for their conversations to be recorded) and trust in their manufacturers and hypothetical third parties (with whom the recordings might be shared/leaked). The interviewer asked whether individuals have read the news or heard anecdotes about unexpected or undesirable behaviors by the smart speakers. These questions created the appropriate context to discuss privacy-preserving probes; while our follow-up questions are capable of biasing the participants, we believe that they are essential in creating the right environment to discuss the ambiguous space of privacy issues surrounding smart speakers.

2. Interaction with Probes: In a randomly generated order, the interviewer briefly introduced the probes and explained their capabilities to the participants. The participants were given time to familiarize themselves with the probe and set it up (*i.e.,* reorganize their existing layout, if needed, to find a suitable location to utilize the probe). If this was not possible in the room where the interview was occurring, the interviewer and participants discussed why this was the case and moved to a more convenient location with a smart speaker, should one exist. By setting up the probe themselves, we expected the participants to gain greater familiarity with its operation and various other nuances (which we discuss later). The random ordering of probes across participants helped to reduce ordering effects. In settings with families, the interviewer asked different family members to interact with the probe individually (in the presence of other members). After setting up the probe, the interviewer asked the participant to issue voice commands after engaging/disengaging the probe. At each step, the interviewer probed the participant about their level of comfort with the probe and how it impacts the usability of their smart speaker. The participants were encouraged to envision future use-cases for each probe and stress-test the probe's functionality. After the interaction with each probe, the interviewer inquired about the participant's level of trust in the probe. Based on the nature of the response, the interviewer asked several follow-up questions to determine reasons for high/low levels of trust. The interviewer proceeded to discuss perceived privacy control, trust level, convenience, and aesthetics of the probes. On average, users interacted with each probe for approximately 20 minutes[4]. These interview questions were designed to elicit critical reflections – the primary aim of the tech probe study.

3. Concluding Discussions: The interviewer engaged the participants in an open-ended discussion about the probes and their impact on their privacy. The interviewer allowed the participants to hold and observe our probes before answering any other questions related to the study. Finally, the interviewer compensated the participants for their time and effort.

We recorded the interviews, resulting in over 30 hours of recordings, and took photographs of (a) the probes in action and (b) areas where the smart speaker is typically used. We then transcribed, coded, and analyzed the interviews using a Grounded Theory approach [7, 14]. The coding was performed with two coders working independently. Our coders were in moderate agreement, with a Cohen's Kappa (κ) of 0.57 [43]. We started the analysis with an open-coding stage to identify more than 200 informal codes that define critical phenomena in the interview transcripts. Using these informal codes, we extracted recurrent themes within the transcripts and converged on a set of 88 formal codes. We further refined the formal codes into 15 axial codes. We organized the codes into three major themes as summarized in Table 3. We believe that the value of the agreement is acceptable for our study, based on previous research [29]. Following common practices in qualitative coding [3], disagreements were discussed by the coders, followed by code reconciliation, resulting in an updated codebook.

4 Observations

In this section, we discuss the central themes that emerged from our analysis. In summary, we found that: (1) participants were reluctant in sacrificing the convenience associated

[4]From our experience, the users were able to familiarize themselves with the mode of operation and installation of these probes in this timeframe.

Attitudes towards Smart Speaker
Characterizes the user's (a) nature and awareness, (b) technological know-how, and (c) trust in smart speaker manufacturer.
Attitudes towards Probe
Characterizes the user's (a) interaction preference, (b) comfort-levels with regards to usage, (c) long-term technological preferences, (d) trust attitude towards probe, and (e) aesthetics and physical footprint preference.
Utility of Probes
Characterizes the user's preference with respect to probe's (a) multi-functionality, (b) cost, (c) ability to provide fine-grained control, and (d) mode of operation *i.e.,* proactive vs. reactive,

Table 3: Summary of the extracted themes.

with smart speakers; hands-free interaction was most preferred, and physical interaction was seen as being not ideal; (2) participants expected bolt-on interventions with existing household decor and to offer cues informing them of the state of both the probe and the smart speaker; and (3) participants had a preference for multi-functionality and fine-grained control (per-user and per-device). Several of the observations we make have been reported earlier [1, 20, 26, 27]. Our work reaffirms them and shows that the sample used for the rest of the analysis reveals consistent perceptions as previous work[5].

4.1 User Attitudes regarding Smart Speakers

1. Types of Users: Through our study, we identified two types of users: (a) *casual users* who utilize their smart speakers for setting alarms, asking questions, etc., and (b) *power users* who have integrated the smart speakers with other devices in their homes (such as smart lights, house monitoring systems, etc.). We also observed that most participants in our first interview phase were casual users, and a majority of those in the second phase were power users. This phenomenon could be based on the pervasive availability of various smart home devices. We observed that power users (and those in the second phase) were also more familiar with passive privacy violations and with the potential for active violations. We observed that power users were more willing to adapt privacy-preserving interventions as their households were more tightly integrated with the smart speaker. We also observed that a majority of the participants did not change their conversations around the smart speakers, but a small minority reported feeling conscious of having discussions around them. Similar observations were made in recent works studying the privacy perceptions/attitudes of smart speaker users [1, 20, 26, 27].

[5]These findings resulted from our observations in 2018, predating many of the works cited here.

2. Understanding of Smart Speaker Operation: A minority of the participants was unaware of how smart speaker's operate, *i.e.,* they were unaware that their voice commands were processed off-site. Participant P_5, for example, believed that the smart speakers did *"some local learning but also some more... I think at some point people were involved in [the processing]... I think there's an automated learning that occurs to adjust itself to the household, right?"* Abdi *et al.* reported similar observations about users having incomplete mental models of the smart home personal assistants [1].

3. Trust in Device Manufacturers: Our participant pool includes fractions (a) that believed that these organizations could be trusted, (b) that believed that some manufacturers were not in the business of collecting personal information and can be trusted, (c) that trusted the manufacturers, but believed that any information collected could be leaked, and (d) that trusted the manufacturers as long as there is personal utility gained from disclosing said information. A recurrent theme was participants' comfort in being recorded because they believed they were part of a large pool of smart speaker users. Participant P_{10} explains, *"I mean we're not planning any nefarious capers... like we're very boring people and therefore nothing that we're talking about would be of interest to anyone on the other end of [the smart speaker]."* Other studies have also studied user's trust in the device manufacturers and have reached similar conclusions [20, 26].

4.2 User Attitudes Regardingt Probes

1. State of Operation: Participants believed that the current designs of the probes make it too inconvenient to use the smart speaker. They state that using them makes the interaction with a smart speaker a two-phase procedure: first, check the state of the probe (engaged vs. not) and disengage if necessary, and then interact with the smart speaker. Some participants stated that the probes added a *mental burden* in terms of remembering its state. Participant P_9 said: *"when you were to power it off say how do you distinguish that state [when it has no power when using PowerCut] from a wake word doing nothing, like I don't know I unmute this right now ... it looks the same."*

2. Ergonomics: Participants were comfortable with the *usability* of the probes. They were easy to set up and use, and the time taken for the probes to activate is acceptable (almost instantaneous in all cases). However, participants expressed dissatisfaction at the longer boot-up times induced by PowerCut. For example, P_8 stated,*"I would find it especially irritating."* Participants suggested that technologies such as Obfuscator that, when disengaged, make the smart speaker*immediately* available were *ideal*. Some participants were concerned about the generalizability of Obfuscator. They believed that the technology is specific to their smart speakers, and would not extend to future smart speakers or smart speakers made by other vendors. Participants were comfortable using a remote

Figure 5: The placement of an Amazon Echo Dot inside an owl-shaped holder in one of the households.

control but felt that their homes have many remotes that could be easily misplaced. When proposing the addition of another remote, P_5 exclaimed, *"they're all over the place, so many remotes! We can't have another remote."* Some participants suggested moving intervention control to a mobile phone app.

3. Trust in Bolt-On Interventions: Finally, participants trust our bolt-on probes more than the built-in mute button. However, participants suggested that trust in a bolt-on intervention would be low if it came from the device manufacturer or any organization that had a similar business model. Participant P_{13} recommended *"a competing company or just a general company that seems like they're like honest"* could develop the interventions. Participants suggested that bolt-on interventions were easier to debug and were easier to understand. However, participants feel that purchasing one bolt-on intervention for every smart speaker would be expensive.

4. Physical Footprint: Participants were concerned with the physical footprint of our probes. While smart speakers were electronic devices, participants often associate them with decorative items (Figure 5) and invest effort in determining where these devices should be placed. A common example of a description about the Obfuscator solution we received was P_2's description: *"a piling on of devices."* Some participants found it difficult to reorganize other items around the smart speaker to facilitate the probe. Additionally, some participants prefer to conceal their outlets, and PowerCut-like interventions would be inconvenient in such scenarios. Participants were uncomfortable with interventions that involve additional wires (as in the case of Obfuscator). Similar observations were made by Pateman *et al.* [35] in the context of the adoption of wearable devices.

4.3 Utility of the Probes

1. Damage to the Environment: Participants were concerned that Obfuscator would cause harm to nearby animals; questions we received upon presenting the Obfuscator were often like P_2's, *"is [this] going to ... make my dog crazy?"* While we did not observe any agitation/discomfort, the participant suggested that their pets could perceive the ultrasound signals and were not bothered. Additionally, participants were concerned about exposing their smart speakers to ultrasound for a prolonged period of time[6].

2. Cost and Multi-Functionality: Cost was repeatedly discussed; participants suggested that the cost of the interventions should not exceed the cost of the smart speaker. Some participants received their smart speakers as gifts. Consequently, they were unable to establish a value for an intervention; P_6 states, *"that's a really interesting question in the sense that I didn't pay for this in the first place. Maybe that's also another reason that I don't have much investment in using this in general."* On the other end of the investment spectrum, we observed that participants who owned multiple smart devices were invested in safeguarding their privacy and were willing to adopt interventions independent of the cost. Participant P_6, who had previously stopped using their smart speakers due to privacy concerns, even stated that they would consider using their device once more given that the interventions were *"cheap... I think would have to rival that remote plug-in cost right because ... it has to be like a cheap utility ... or a cheap accessory like that."* Obfuscator could be used in a proactive way *i.e.,* always-on, or in a reactive way *i.e.,* use when needed. Participants felt that a reactive approach, though tedious, would be easier to understand. Participants also believed that cost could be justified if the intervention provided multiple features. This could be achieved by integrating the design of Obfuscator with other home decors, such as lamps, lights, clocks, radios.

3. Multi-user and Multi-device Environments: The final observation we make is an extension to multi-user and multi-device environments; we observed that in some households, some participants preferred to utilize the intervention more than others. Also, different types of users might exhibit different privacy requirements when interacting with smart speakers. In such scenarios, they desired customized usage profiles based on their requirements *i.e., access control per-user*. Recent research has also indicated the need for access control flexibility in multi-user smart homes [47]. Another observation we make is that some participants preferred to have one intervention (like Obfuscator) being used to preserve privacy against a wide range of smart speaker-like devices. In such scenarios, *access control per-device* was desired. Based on the current design of the Obfuscator prototype, meeting both these requirements is challenging and requires further research. One research direction to make access control per-user more feasible is establishing default privacy options depending on the expected user privacy profiles [2], such as owner vs. visitor.

Concrete Recommendations
1. Aesthetics: The interventions should be offered in different forms, shapes, and colors to fit within people's decors and furniture.
2. Physical Footprint: The footprint of the intervention should be small enough to not force a reorganization of the layout of the owner's house.
3. Multi-Functionality: The intervention is better when providing additional functionality (such as a clock) to reduce its footprint and integrate better with home decor.
4. Ease of Deployment and Understanding: Battery-powered interventions are easier to deploy.
5. Ease of Understanding: A proper understanding of the privacy metaphor improves the adoption of interventions.
6. Trust in Technology: Trustworthy interventions are bolt-on, not network connected, designed by a different trustworthy organization, and pose no additional risk.
7. Mode of Interaction: Using the intervention should not change the interaction with the smart speaker. Hands-free interaction is preferable.
8. Informative Cues: Interventions should offer cues that communicate their state. Visual, auditory, or text cues might be applicable depending on the deployment.
9. Cost: The intervention should cost less than the smart speaker.
10. Fine-grained Privacy Control: The intervention can offer per-user and per-smart speaker privacy controls.
11. Awareness: Awareness of privacy violations increases trust in intervention designers.

Table 4: Summary of the identified design guidelines.

5 Design Implications

Based on the findings from § 4, we make concrete recommendations (based on our findings) on how to design privacy-preserving interventions. The design recommendations are along axes specified in Table 4.

1. Aesthetics: We observed the aesthetics of the privacy interventions to be an important issue for our participants. Participants preferred the interventions to match their individual decorating styles (one example is shown in Figure 5). Many participants suggested that the intervention should come in different forms, shapes, and colors, enabling easier integration within their home decor. As individual tastes vary widely, devising a one-fits-all design is challenging. *One possible approach is to explore different design options for different types of users, including shapes, forms, colors, and material*. This approach has been successful with smart speakers, where participants feel comfortable with the aesthetic of the smart speaker. For example, Amazon has four variants of their Echo featuring combinations of forms and fabric colors.

2. Physical Footprint: Since the smart speakers we considered were small and compact, participants preferred a similar physical footprint for the interventions. Participants expressed concerns regarding the size of both PowerCut and Obfuscator, enquiring if a similar functionality could be achieved with a smaller probe. They believed that using Obfuscator (which needs to be proximate to the smart speaker) requires them to significantly reorganize their existing home decor layout. While the form factor of PowerCut can be reduced trivially, doing so for Obfuscator is challenging; the size and shape of the horn speaker in our current probe were chosen to ensure maximum ultrasound distribution and coverage. Extending such a design to (newer) smart speakers that are larger, or have a different orientation for the microphone inputs, will require rethinking the design and form factor. In summary, interventions that require proximity to the smart speaker need to be designed such that their form factor is comparable to the smart speaker. To achieve such a design, *one recommendation is to design the Obfuscator-style intervention as a stand (upon which the smart speaker can be placed), or as an artifact that can be placed above the smart speaker.* In both designs, the intervention will generate a veil of ultrasound around the entire smart speaker (similar to the horn speaker case that we had designed and evaluated).

3. Multi-Functionality: Closely tied to the aesthetics, participants indicated preference toward an intervention (specifically Obfuscator) that offered features beyond privacy-preservation. They suggested that the Obfuscator intervention could be combined with other household artifacts, such as a lamp, radio, clock, which would further improve adoption. Additionally, *multi-functionality provides an alternative avenue for customizing the probe, making it easier to integrate with existing household decoration*. Such products alleviate the social stigma of being labeled as overly privacy-conscious; such stigma is another reason why the adoption of privacy-preserving interventions is currently low.

4. Ease of Deployment: Participants state that they prefer having a solution that is easy to deploy in their homes; the biggest impediment to any intervention similar to PowerCut is its requirement for an outlet. Many participants preferred to conceal the interventions' wiring, and the nearby outlets can be hard to reach. Attaching PowerCut to wall outlets, even once, requires considerable re-positioning of other devices and their wires. Attaching Obfuscator would require an additional outlet, which is not always readily available. One naïve solution would be to split the outlet among multiple devices. Participants suggested that an Obfuscator design capable of operating on batteries would be more preferred, even if this required periodic replacement.

[6] A detailed study is needed to understand the impact of ultrasound on electronic devices.

Recommendation: Combining the above four observations, we recommend designing interventions in one of two forms: (a) a stand to hold the smart speaker, or (b) a sleeve for the smart speaker (refer Figures 7 and 6). Based on some preliminary analysis, we observe that there is a demand for such artifacts based on our analysis of reviews for such products, and we believe such designs would promote adoption. Since the intervention is not operational in an *always-on* mode, it may be battery-powered — doing away with the requirement for an outlet.

5. Ease of Understanding: All participants were able to easily grasp the metaphor associated with PowerCut, but the technology behind Obfuscator proved complicated for some; some users were unfamiliar with how ultrasound induces a deafening effect. Thus, interventions whose operation is easy to explain may be preferred. This is particularly the case because, while Obfuscator is easy to use once deployed, debugging it may pose problems for users who lack a proper understanding of its operation. We also believe that understanding the detriments (if any) of ultrasound towards humans, animals, and other electronics may put users at ease.

6. Trust in the Technology: Participants were more comfortable with technologies that they believe will survive the "test of time," *i.e.,* be useful for smart speaker models in the future. As discussed earlier, trust also stems from knowing that the interventions do not pose any additional risk. Specifically, we observed that (a) participants wanted to know about any detriments introduced by the interventions, such as potential damage to the smart speaker by frequently disconnecting it from its power source or subjecting it to ultrasound; and (b) our current interventions are not network connected and do not present the same risks as the smart speakers. Finally, participants preferred our bolt-on interventions as opposed to the built-in interventions as they were designed by an organization they trusted (more than the smart speaker manufacturers).

Recommendation: Combining the two points stated above, a concrete design recommendation is to communicate the science behind the operation of the PowerCut-style intervention with a more relatable metaphor or through an interactive demonstration of the intervention's operation. By doing so, we are able to provide more intuition on failure scenarios, which can enable more efficient debugging. This process also assuages any fears related to smart speaker damage or possible harm to nearby entities (such as pets).

7. Mode of Interaction: We observed that participants placed a high value on the *convenience* of using smart speakers, which they are not willing to compromise. Thus, interventions that, when engaged, delay the smart speaker operation (as in the case of PowerCut) are not preferred (even though PowerCut provably preserves privacy, and its mode of operation is very easy to understand). Additionally, any form of physical interaction, be it using remotes or buttons, is far from ideal; some participants expressed preference toward using an app on their smartphones.

Recommendation: We believe that future interventions must be designed so as to have minimal disruption to the convenience of the use of these systems. An ideal design would have a voice interface that allows the user to control it as they control their smart speakers. However, such an always-on and listening privacy-preserving solution can have the same pitfalls as smart speakers, and they must be designed in a manner that does not erode user trust; the mechanism to provide privacy (via a voice-interface) must not become a mechanism for exfiltrating sensitive user conversation (as such a mechanism may require to be network connected). For example, they can lack a network interface to provide the users with hard privacy assurances. Another issue that may arise with voice-activated interventions is erroneous activations; understanding how this can be minimized requires additional research.

8. Informative Cues: As stated earlier, some participants concealed their smart speakers and would prefer concealing their interventions as well. Some participants take this notion to the extreme; they believe that any electronic device that does not provide extensive visual information should be concealed. Thus, visual cues are not ideal in all situations. Additionally, participants suggested that the *red* light on the PowerCut intervention suggested that the intervention was broken, as opposed to indicating the state of the intervention. Interacting with the smart speaker when the intervention is enabled helps users determine the state of the smart speaker (operational vs. not), but such an approach is reactive. Participants indicated a preference for a *proactive* approach.

Recommendation: We propose two recommendations for such settings: (a) the state of the intervention (i.e., engaged vs. disengaged) by communicating to a device that is more optimally placed for being viewed (such as a TV) — this can be done using some form of a closed network connection between the TV and the device via Bluetooth, or (b) the intervention provide auditory cues, where the Obfuscator-style intervention can announce using speech or text that the smart speaker is inactive when users try to activate it.

9. Cost: Another factor that impacts adoption is the cost of the smart speaker. A large fraction of our participants owns smart plugs similar to PowerCut, leading us to believe that such an

intervention is affordable. However, the cost of prototyping Obfuscator was $70, exceeding the cost of smart speakers (priced at approx. $30). This cost includes the price of the commodity parts needed to construct the probe. Participants believe that the cost of the intervention should not exceed the cost of the smart speaker; this is especially true if the intervention can provide privacy protection against a single smart speaker. We believe that if such an intervention would be adopted widely, the production costs could be amortized (and thus have no concrete recommendation to make with regards to minimizing cost). Additionally, understanding the engineering requirements to design an Obfuscator-like intervention that provides privacy against various smart speakers located at different parts of a home requires independent research.

10. Fine-grained Privacy Control: Several households owned more than a single smart speaker, and they had members with different (and potentially conflicting) privacy requirements. Thus, we believe that there is a requirement for (a) fine-grained control *per user*, and (b) fine-grained control *per smart speaker*. For the latter, a naïve solution would be to deploy one intervention per smart speaker, but depending on the cost per intervention, such a solution may not scale. Providing per-user control is a more challenging problem; it requires understanding how disparate the privacy requirements are, how frequently users are utilizing a smart speaker together, and how to mitigate conflicts should they arise.

> **Recommendation:** An ideal design would provide privacy protection for more than one smart speaker. This design could be conceptualized as smaller interventions co-located with the smart speakers but controlled centrally (through some form of closed network).

11. Effect of Awareness: Based on our interview questions, we observed the following trend amidst the participants of our interview phases: participants of our second phase are more concerned about the potential privacy threats from the smart speakers (in comparison to the participants of the first phase, who are also concerned). This concern stems from increased awareness, recent smart speaker mishaps, erroneous code used in them, and immoral practices by device manufacturers. Based on our discussion, we observed that participants believe that these issues are not being seriously audited by the device manufacturers. Discussing various loopholes that can be implemented in the built-in interventions in the status quo (*i.e.,* local wake word processing and the mute button) also increased participants' awareness.

5.1 Consolidated Recommendation

We consolidate the design recommendations based on the aforementioned discussion and provide concrete design guidelines.

Aesthetics, Utility, and Accessibility: Obfuscator-like interventions should be incorporated in accessories that users are already adopting, such as stands and holders (the "owl" shown in Fig 5). Since the completion of our work, we have seen an emergence of a market for such accessories. Additionally, the jamming device should be hidden within a device that is multi-functional, privacy being the secondary functionality. Further, the jamming device should be always-on; the user can access the smart speaker through hands-free interactions, such as gesture-based interaction through wireless sensing. A chime can be played to indicate that the smart speaker is currently active.

Cost & Centralized Control: Obfuscator-like jamming systems rely on directionality to enable their functionality. Thus, it is unclear if there can be *one* of such solutions for *multiple* smart speakers in a home environment. However, many such interventions can be controlled through a centralized interface, such as a single remote control or mobile phone app. Future research is required to better understand the requirements of such a control interface and design it.

Building User Trust: To enhance trust in such interventions, video (or other forms of) presentations/materials can be made to indicate that current smart speakers are purported to exfiltrate home conversation through the use of the public internet. Once this is established, we can educate users of the fact that Obfuscator-like interventions are not connected to any network and consequently can not share sensitive (or any other) information. To further strengthen user belief in bolt-on solutions, end-users can be educated about issues with built-in solutions. Notably, make changes to any built-in solution after deployment requires device manufacturers to regularly and reliably share software updates. However, installing such updates is a challenging proposition to even tech-savvy users [36]. Additionally, as the ecosystem of such smart speaker devices is fast evolving, manufacturers will often not provide support to (a large volume of) smart speakers that were deployed in the past [39]. Additionally, information about software updates (needed for built-in solutions) is not easily accessible, resulting in periods of privacy loss [17].

Accessibility in a Multi-User Environment: Since different users may have different privacy requirements, interventions may be designed to operate with different profiles, such as always-on versus selectively turned on. However, choosing the profile may require (a) explicit user interaction with the intervention, which may be inconvenient, or (b) using auxiliary hardware to identify the users [12].

Enhancing Awareness: Finally, we recommend an on-boarding process that educates the users about the potential privacy threats from accidental/malicious activations, the technology underlying the operation of the intervention, and how to utilize the privacy controls. Such an on-boarding process can take place through voice prompts or an external app; it will increase the user's awareness of the privacy issues as well as improve the user's trust in the probe.

6 Related Work

Privacy Perceptions: The methodology of our study is most similar to Zheng *et al.* [50], and Kaaz *et al.* [23]. They attempt to understand the privacy perceptions of users living in homes with various IoT devices. Similar to our work, surveys are carried out in [4, 9, 28], where the authors try to identify the various challenges associated with setting up and using these devices. Zeng *et al.* [46] and Lau *et al.* [26] study smart speaker users' reasons for adoption through a combination of a detailed diary study and in-home interviews. Similar to this work, we observe that smart speaker users are not privacy-conscious because of the lack of value they associate with their conversational data. Along a similar vein, Abdi *et al.* [1] find that users have incomplete mental models of smart speakers. Similar to our work, they use this understanding to present design recommendations. Some of our findings are coherent with those of Malkin *et al.* [27], *e.g.,* participants are unaware that their conversations are being recorded and stored.

We stress that the primary contribution of our work is *not* in ascertaining the privacy perceptions people have about smart speakers (as done in earlier studies). We wish to understand users' perceptions towards privacy-enhancing technologies and to use this insight to guide the design of both smart speakers and such technologies.

Probe Design: Prior research has investigated system-level solutions to these privacy threats. Feng *et al.* [12] propose continuous authentication as a mechanism to thwart privacy issues related to smart speakers. In our previous work, we propose using ultrasound jamming to address stealthy recording *et al.* [13]. In this work, we wish to validate the usability claims made by the above; consequently, we base our intervention design on the above proposals. The works of McMillan *et al.* [30] and Mhaidli *et al.* [31] provide hands-free alternatives. However, the introduction of a camera to measure gaze introduces privacy concerns. This also requires the user to be in the line of sight of the smart speaker, which reduces its usability. Solutions based on pitch and volume [31] also suffer from similar proximity issues and fail to eliminate privacy violations due to accidental activations.

The Alias project [24] is designed to achieve similar goals to ours. This solution constantly plays noise through a small speaker placed atop the smart speaker and stops the noise upon hearing a custom wake word. Their solution differs from Obfuscator in two ways. First, the Alias intervention does not use ultrasound; the reduced form factor is achieved by not using horn speakers, which are crucial for transmitting ultrasound. Second, the Alias intervention obscures the visual cue provided by the smart speaker; such a design is not preferred. Similarly, work by Chen *et al.* [8] designs a wearable intervention. Wearable solutions offer support in some scenarios, *e.g.,* mobile situations. However, they offer poor support for smart speaker due to lack of proximity to the device. Our experiments with ultrasound-based jamming revealed that the direction of the jamming device and the distance to the smart speaker impact its performance. Additionally, Chen *et al.* do not evaluate the user-related aspects of the intervention, such as user acceptance, aesthetics, and trust.

Design Studies: To safeguard privacy and security in the smart home, Zeng *et al.* [47] prototyped a smart home app and evaluated its effectiveness through a month-long in-home user study with seven households; the users are assumed to be non-adversarial and cooperative. They used their findings to guide future designs for smart home applications. To achieve similar goals as ours, but for smart homes (as opposed to smart speakers), Yao *et al.* [45] adopted a co-design approach and designed solutions with non-expert users. We borrow our study methodology from the work of Odom *et al.* [33]; technology probe studies serve multiple purposes related to designing, prototyping, and field testing the interventions.

7 Conclusions

We presented the design and prototyping of two privacy-preserving interventions: 'Obfuscator' targeted at disabling recording at the microphones, and 'PowerCut' targeted at disabling power to the smart speaker. We presented our findings from a technology probe study involving 24 households that interacted with our prototypes, aimed to gain a better understanding of this design space. Our study revealed several design dimensions for the design of privacy interventions for smart speakers, including multi-functionality, trustworthiness, cues, interaction mode, and ease of deployment.

Acknowledgments

We would like to thank Christopher Little and Thomas Linden who helped with the interviews. We would also like to thank Mariam Fawaz who assisted with the photographs of the interventions, and Yilong Li who assisted with the fabrication of Obfuscator. Finally, we would like to thank the anonymous reviewers and our shepherd for their constructive feedback. Varun, Suman, and Kassem were supported in part through the following US NSF grants: CNS-1838733, CNS-1719336, CNS-1647152, CNS-1629833, CNS-1942014, and CNS-2003129 and an award from the US Department of Commerce with award number 70NANB21H043.

References

[1] Noura Abdi, Kopo M Ramokapane, and Jose M Such. More than smart speakers: Security and privacy perceptions of smart home personal assistants. In *Symposium on Usable Privacy and Security (SOUPS)*, 2019.

[2] Noura Abdi, Xiao Zhan, Kopo M Ramokapane, and Jose Such. Privacy norms for smart home personal assistants. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2021.

[3] Jyoti Belur, Lisa Tompson, Amy Thornton, and Miranda Simon. Interrater reliability in systematic review methodology: exploring variation in coder decision-making. *Sociological methods & research*, 50(2):837–865, 2021.

[4] AJ Brush, Bongshin Lee, Ratul Mahajan, Sharad Agarwal, Stefan Saroiu, and Colin Dixon. Home automation in the wild: challenges and opportunities. In *proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2115–2124. ACM, 2011.

[5] Marion Buchenau and Jane Fulton Suri. Experience prototyping. In *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques*, pages 424–433. ACM, 2000.

[6] Nicholas Carlini, Pratyush Mishra, Tavish Vaidya, Yuankai Zhang, Micah Sherr, Clay Shields, David Wagner, and Wenchao Zhou. Hidden voice commands. In *25th USENIX Security Symposium (USENIX Security 16)*, pages 513–530, Austin, TX, August 2016. USENIX Association.

[7] Kathy Charmaz and Liska Belgrave. Qualitative interviewing and grounded theory analysis. *The SAGE handbook of interview research: The complexity of the craft*, 2:347–365, 2012.

[8] 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*, CHI '20, 2020.

[9] Eun Kyoung Choe, Sunny Consolvo, Jaeyeon Jung, Beverly Harrison, and Julie A Kientz. Living in a glass house: a survey of private moments in the home. In *Proceedings of the 13th international conference on Ubiquitous computing*, pages 41–44. ACM, 2011.

[10] CNN. Google admits its new smart speaker was eavesdropping on users. https://web.archive.org/web/20210226070734/https://money.cnn.com/2017/10/11/technology/google-home-mini-security-flaw/index.html, 2017.

[11] 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.

[12] Huan Feng, Kassem Fawaz, and Kang G Shin. Continuous authentication for voice assistants. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, pages 343–355. ACM, 2017.

[13] Chuhan Gao, Varun Chandrasekaran, Kassem Fawaz, and Suman Banerjee. Traversing the quagmire that is privacy in your smart home. IoT S&P '18, page 22–28, New York, NY, USA, 2018. Association for Computing Machinery.

[14] Barney G Glaser and Anselm L Strauss. *Discovery of grounded theory: Strategies for qualitative research*. Routledge, 2017.

[15] Google Home Help. [Fixed issue] Google Home Mini touch controls behaving incorrectly. https://support.google.com/googlehome/answer/7550221?hl=en, 2018.

[16] The Guardian. Apple contractors 'regularly hear confidential details' on Siri recordings. https://web.archive.org/web/20210513003110/https://www.theguardian.com/technology/2019/jul/26/apple-contractors-regularly-hear-confidential-details-on-siri-recordings, 2019.

[17] Kashmir Hill. 'baby monitor hack' could happen to 40,000 other foscam users. https://web.archive.org/web/20210527231505/https://www.forbes.com/sites/kashmirhill/2013/08/27/baby-monitor-hack-could-happen-to-40000-other-foscam-users/?sh=352cfe4558b5, 2013.

[18] Kashmir Hill and Surya Mattu. The house that spied on me. http://web.archive.org/web/20210518035909/https://gizmodo.com/the-house-that-spied-on-me-1822429852, 2018.

[19] Gary Horcher. Woman says her amazon device recorded private conversation, sent it out to random contact. http://web.archive.org/web/20210412111205/https://www.kiro7.com/news/local/woman-says-her-amazon-device-recorded-private-conversation-sent-it-out-to-random-contact/755507974/, 2018.

[20] 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.

[21] Hilary Hutchinson, Wendy Mackay, Bo Westerlund, Benjamin B. Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, Nicolas Roussel, and Björn Eiderbäck. Technology probes: Inspiring design for and with families. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 17–24, New York, NY, USA, 2003. ACM.

[22] Haojian Jin, Minyi Liu, Kevan Dodhia, Yuanchun Li, Gaurav Srivastava, Matthew Fredrikson, Yuvraj Agarwal, and Jason I Hong. Why are they collecting my data?: Inferring the purposes of network traffic in mobile apps. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(4):173, 2018.

[23] Kim J. Kaaz, Alex Hoffer, Mahsa Saeidi, Anita Sarma, and Rakesh B. Bobba. Understanding user perceptions of privacy, and configuration challenges in home automation. In *Visual Languages and Human-Centric Computing (VL/HCC), 2017 IEEE Symposium on*, pages 297–301. IEEE, 2017.

[24] Bjorn Karmann. Project alias. http://bjoernkarmann.dk/project_alias, 2018.

[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, Baltimore, MD, August 2018. USENIX Association.

[26] 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 Conference on Computer-Supported Cooperative Work and Social Computing (CSCW)*, 2018.

[27] 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.

[28] Faith McCreary, Alexandra Zafiroglu, and Heather Patterson. The contextual complexity of privacy in smart homes and smart buildings. In *International Conference on HCI in Business, Government and Organizations*, pages 67–78. Springer, 2016.

[29] Mary L McHugh. Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3):276–282, 2012.

[30] Donald McMillan, Barry Brown, Ikkaku Kawaguchi, Razan Jaber, Jordi Solsona Belenguer, and Hideaki Kuzuoka. Designing with gaze: Tama–a gaze activated smart-speaker. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–26, 2019.

[31] 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.

[32] Assaf Hurwitz Michaely, Xuedong Zhang, Gabor Simko, Carolina Parada, and Petar Aleksic. Keyword spotting for google assistant using contextual speech recognition. In *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pages 272–278. IEEE, 2017.

[33] William Odom, Richard Banks, David Kirk, Richard Harper, Siân Lindley, and Abigail Sellen. Technology heirlooms?: Considerations for passing down and inheriting digital materials. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 337–346, New York, NY, USA, 2012. ACM.

[34] Danny Palmer. Amazon's Alexa could be tricked into snooping on users, say security researchers. https://web.archive.org/web/20210301140435/https://www.zdnet.com/article/amazons-alexa-could-be-tricked-into-snooping-on-users-say-security-researchers/, 2018.

[35] Matthew Pateman, Daniel Harrison, Paul Marshall, and Marta E Cecchinato. The role of aesthetics and design: wearables in situ. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–6, 2018.

[36] John Pescatore. Securing the "internet of things" survey. https://www.sans.org/reading-room/whitepapers/covert/paper/34785, 2014.

[37] Nirupam Roy, Haitham Hassanieh, and Romit Roy Choudhury. Backdoor: Sounds that a microphone can record, but that humans can't hear. *GetMobile: Mobile Computing and Communications*, 21(4):25–29, 2018.

[38] Nirupam Roy, Sheng Shen, Haitham Hassanieh, and Romit Roy Choudhury. Inaudible voice commands: The long-range attack and defense. In *15th USENIX Symposium on Networked Systems Design and Implementation (NSDI 18)*, pages 547–560. USENIX Association, 2018.

[39] Bruce Scheier. The internet of things is wildly insecure — and often unpatchable. http://web.archive.org/web/20210520212158/https://www.wired.com/2014/01/theres-no-good-way-to-patch-the-internet-of-things-and-thats-a-huge-problem/, 2014.

[40] 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.

[41] Siddharth Sigtia, Rob Haynes, Hywel Richards, Erik Marchi, and John Bridle. Efficient voice trigger detection for low resource hardware. In *Interspeech*, pages 2092–2096, 2018.

[42] 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*, SenSys '20, page 298–311, New York, NY, USA, 2020. Association for Computing Machinery.

[43] Anthony J Viera, Joanne M Garrett, et al. Understanding interobserver agreement: the kappa statistic. *Fam med*, 37(5):360–363, 2005.

[44] voicebot.ai. Nearly 90 million u.s. adults have smart speakers, adoption now exceeds one-third of consumers. https://web.archive.org/web/20210503050256/https://voicebot.ai/2020/04/28/nearly-90-million-u-s-adults-have-smart-speakers-adoption-now-exceeds-one-third-of-consumers/, 2020.

[45] 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*, page 198. ACM, 2019.

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

[47] Eric Zeng and Franziska Roesner. Understanding and improving security and privacy in multi-user smart homes: A design exploration and in-home user study. In *28th USENIX Security Symposium (USENIX Security 19)*, pages 159–176, Santa Clara, CA, August 2019. USENIX Association.

[48] Guoming Zhang, Chen Yan, Xiaoyu Ji, Tianchen Zhang, Taimin Zhang, and Wenyuan Xu. Dolphinattack: Inaudible voice commands. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 103–117. ACM, 2017.

[49] Nan Zhang, Xianghang Mi, Xuan Feng, XiaoFeng Wang, Yuan Tian, and Feng Qian. Dangerous skills: Understanding and mitigating security risks of voice-controlled third-party functions on virtual personal assistant systems. In *2019 IEEE Symposium on Security and Privacy (SP)*, pages 1381–1396, 2019.

[50] Serena Zheng, Noah Apthorpe, Marshini Chetty, and Nick Feamster. User perceptions of smart home IoT privacy. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW), November 2018.

Appendix

A Formal Codes

1. privacy awareness vs education = awareness function of education
2. privacy awareness vs education = not equivalent
3. user technical knowledge = high
4. user technical knowledge = medium
5. user technical knowledge = low
6. user technical knowledge = varies in home
7. user education level = high
8. user education level = medium
9. user education level = low
10. user has concern = yes listening
11. user has concern = yes recording
12. user has concern = yes other
13. user has concern = no
14. user trust large orgs = yes
15. user trust large orgs = case by case
16. user trust large orgs = no
17. user trust third party = yes
18. user trust third party = no
19. user type = power user
20. user type = simple user
21. user accepts listening if = choose over recording
22. user accepts listening if = machine only
23. user accepts recording if = utility
24. user solution choice = discard device
25. user solution choice = unplug device
26. user solution choice = mute
27. user solution choice = remote plug
28. user solution choice = obfuscator
29. user believes in intervention = maybe
30. user believes in intervention = no
31. va state listening = wake word only
32. va state listening = yes
33. va state recording = yes
34. va state recording = non human
35. va state issue attribution = bugs
36. va state issue attribution = unaware
37. va state data use = mundane
38. va state data use = nefarious
39. ecosystem factor = space for solution
40. ecosystem factor = utility of visual cues
41. mute aethestic = acceptable
42. mute aethestic = not acceptable
43. mute haptics = acceptable
44. mute haptics = not acceptable
45. mute form = acceptable
46. mute form = not acceptable
47. mute usability = acceptable
48. mute usability = not acceptable
49. mute concern = privacy protection
50. remote plug aethestic = acceptable
51. remote plug aethestic = not acceptable
52. remote plug form = acceptable
53. remote plug form = not acceptable
54. remote plug haptics = acceptable
55. remote plug haptics = not acceptable
56. remote plug usability = acceptable
57. remote plug usability = not acceptable
58. remote plug concern = boot up time
59. obfuscator aethestic = acceptable
60. obfuscator aethestic = not acceptable
61. obfuscator form = acceptable
62. obfuscator form = not acceptable
63. obfuscator haptics = acceptable

64. obfuscator haptics = not acceptable
65. obfuscator usability = acceptable
66. obfuscator usability = unsure
67. obfuscator usability = not acceptable
68. obfuscator concern = animals
69. obfuscator concern = harm device
70. ideal solution interface = voice
71. ideal solution interface = hands free
72. ideal solution interface = app
73. ideal solution integration = built in
74. ideal solution integration = bolt on
75. ideal solution form = minimal
76. ideal solution form = distributed for devices
77. ideal solution aethestic = multifunctional
78. ideal solution haptics = important
79. ideal solution haptics = not important
80. ideal solution ux = minimal interaction frequency
81. ideal solution ux = no downtime
82. ideal solution ux = no single point control
83. ideal solution ux = single point control
84. ideal solution other = all local
85. ideal solution developer = first party
86. ideal solution developer = third party
87. decision factor = cost
88. decision factor = privacy awareness

B Items in the Commercial Market

We provide screenshots of several cases/sleeves used to encase the Amazon Echo smart speaker. Similar products can be found for the Google smart speaker as well.

Figure 6: A case-like enclosing for Amazon Echo, on Amazon.

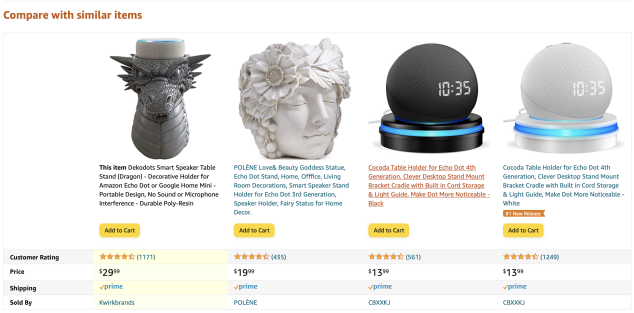


Figure 7: Case-like enclosing recommended by Amazon, for Amazon Echo, on Amazon.