# Tangible Privacy for Smart Voice Assistants: Measuring Users' Attitudes Towards User-Centric Sensor Designs

IMTIAZ AHMAD, Department of Computer Science, Indiana University Bloomington, USA
TASLIMA AKTER, Department of Computer Science, Indiana University Bloomington, USA
ZACHARY BUHER, Department of Computer Science, Indiana University Bloomington, USA
ROSTA FARZAN, School of Computing and Information, University of Pittsburgh, USA
APU KAPADIA, Department of Computer Science, Indiana University Bloomington, USA
ADAM J. LEE, Department of Computer Science, University of Pittsburgh, USA

Smart voice assistants such as Amazon Alexa and Google Home are becoming increasingly pervasive in our homes. Despite their benefits, their miniaturized and embedded sensors (such as cameras and microphones) raise serious privacy concerns related to surveillance and eavesdropping. Recent work on understanding users' privacy concerns around these devices has highlighted the need for 'tangible privacy' where *control* and *feedback* mechanisms can provide a more assured sense of privacy (as with camera lens covers). To address this gap in the design of IoT devices, especially in the case of disabling microphones, we evaluate several designs of smart voice assistants that incorporate (or not) tangible control and feedback mechanisms. By comparing user's perceptions of risk, trust, reliability, usability, and control for these designs in a between-subjects online experiment (N=261), we find that users considered devices with tangible built-in physical controls to be more trustworthy *and* usable than non-tangible mechanisms. Our findings present an approach for tangible, assured privacy especially in the context of embedded microphones.

CCS Concepts: • **Security and privacy** → *Usability in security and privacy*; • **Human-centered computing** → *Empirical studies in ubiquitous and mobile computing*.

Additional Key Words and Phrases: IoT devices, tangible design, privacy assurance, smart voice assistant

**ACM Reference Format:**
Imtiaz Ahmad, Taslima Akter, Zachary Buher, Rosta Farzan, Apu Kapadia, and Adam J. Lee. 2018. Tangible Privacy for Smart Voice Assistants: Measuring Users' Attitudes Towards User-Centric Sensor Designs. 1, 1 (March 2018), 31 pages. https://doi.org/10.1145/1122445.1122456

## 1 INTRODUCTION

Network-connected smart voice assistants (SVA) are becoming increasingly pervasive in our households. These devices increase users' convenience with many daily tasks such as accessing information, setting alarms, playing songs, scheduling events, or controlling smart home appliances

Authors' addresses: Imtiaz Ahmad, Department of Computer Science, Indiana University Bloomington, 700 N Woodlawn Ave, Bloomington, Indiana, USA, imtahmad@iu.edu; Taslima Akter, Department of Computer Science, Indiana University Bloomington, 700 N Woodlawn Ave, Bloomington, Indiana, USA, takter@iu.edu; Zachary Buher, Department of Computer Science, Indiana University Bloomington, 700 N Woodlawn Ave, Bloomington, Indiana, USA, zkbuher@iu.edu; Rosta Farzan, School of Computing and Information, University of Pittsburgh, 135 North Bellefield Avenue, Pittsburgh, Pennsylvania, USA, rfarzan@pitt.edu; Apu Kapadia, Department of Computer Science, Indiana University Bloomington, 700 N Woodlawn Ave, Bloomington, Indiana, USA, kapadia@indiana.edu; Adam J. Lee, Department of Computer Science, University of Pittsburgh, 210 S. Bouquet St., Pittsburgh, Pennsylvania, USA, adamlee@pitt.edu.

XXXX-XXXX/2018/3-ART $15.00
https://doi.org/10.1145/1122445.1122456

(such as light bulbs and coffee makers) by using simple voice commands. According to NPR and Edison Research, by the end of Spring 2020 about 24% of all adult Americans (around 60 million people) owned at least one smart voice assistant.[1] This popularity is only expected to grow; by 2027, the smart voice assistant market is predicted to reach over $19 billion.[2]

Despite their positive role in making our lives convenient, researchers have identified potential privacy threats perceived by the users of these voice assistants [1, 59, 64]. Interestingly, microphones are considered as one of the most invasive and privacy-violating sensors, yet are an integral part of these devices [16, 73]. Moreover, users have been shown to have limited knowledge of how their smart voice assistants collect, store, and process data [1, 64]. Smart voice assistants are designed to operate in an 'always-on' mode, which essentially means the devices are equipped with embedded microphones that are always active and continuously listening for a pre-specified 'wake word', even when the user has no intention of interacting with the devices. Although these devices (e.g., Amazon Echo) are not supposed to record any audio until they detect the wake word [37], recent patent filings by Google and Amazon show the intention to use such devices to listen to words that indicate users' interests (e.g., "I love skiing") and use this information for targeted advertisements [99]. Furthermore, these devices can be accidentally triggered when they misinterpret the wake word causing them to record conversations that were not intended to be recorded [58]. For example, in May 2018, a Portland family found out that a private conversation in their home was recorded by Amazon's Alexa and then sent to a person who is in the family's contact list because the device had misinterpreted their conversation as a sequence of commands [48]. To make things worse, it has been demonstrated that voice assistants can be controlled by voice commands that are unintelligible to human listeners [18]. Similarly, Zhang et al. showed that voice assistants can be tricked to receive and respond to inaudible voice commands. This high-frequency attack, known as a *DolphinAttack*, cannot be perceived by humans [106] making it infeasible for an average user to guard against such privacy vulnerabilities. Although prior works have reported users are willing to trade off their privacy for the convenience and benefits provided by these devices [44, 59], this behavior can be attributed to the little opportunity provided to selectively control the devices' features and the privacy implications [36, 59].

Given the ever-increasing adoption of smart voice assistants coupled with their existing potential privacy vulnerabilities, there is a pressing need to explore design solutions for voice assistants that provide users with convenient and trustworthy privacy controls. In this paper, we aim to explore several dimensions of the design space for privacy-preserving smart voice assistants. In particular, we explore design solutions with hardware-based privacy enhancing mechanisms based on the principles of 'tangible privacy' mechanisms proposed by Ahmad et al [5]. The authors argue for "tangible privacy" mechanisms as a combination of "privacy control and feedback mechanisms that are tangible, i.e., manipulated or perceived by touch, and of high assurance, i.e., they provide clear confidence and certainty of privacy to observers." Devices equipped with tangible privacy mechanisms must have (a) tangible control mechanisms to allow people to unambiguously control the device's data collection with haptic, direct manipulation, which enables users "to grab, feel, and move the important elements" of the control mechanism [5, 49], and (b) high assurance feedback mechanisms that provide a clear and definite sense of awareness of what data is being collected to people in the device's vicinity. This approach provides a "seamless integration of representation and control," and offers a legible mapping between user actions and mediated effects [5, 96, 100]. A primary example of such a mechanism for cameras is to provide lens covers, which have both these properties, where most importantly users in the vicinity 'know' (assured

[1] https://www.nationalpublicmedia.com/insights/reports/smart-audio-report

[2] https://www.researchandmarkets.com/reports/4806139/smart-speaker-global-market-trajectory

feedback) the camera cannot see them. This approach is particularly encouraging because tangible control mechanisms give tactile feedback which is found to improve decision making [78], provide confidence in evaluating system's functionality [77], and enable users to read the system's state from the manipulated object [97]. Similarly, tangible feedback mechanisms are found to successfully raise awareness among users [85, 107], improving the understanding of the system's states [96].

Despite the potential for tangible privacy mechanisms to provide higher levels of privacy assurance to users, more research is needed to evaluate the effectiveness of tangible control and feedback mechanisms in providing users with high assurance of their privacy, especially in the case of embedded microphones. More recently, companies like Facebook[3] and Apple[4] have recognized the need for hardware based microphone disconnects (with the use of mute button) for their devices. However, more work is needed to better design such mute buttons to convince users that there is indeed a hardware disconnect as suggested by prior work that users doubt the effectiveness of mute buttons and prefer unplugging devices entirely to ensure their privacy [5, 59]. Thus, we seek to understand whether suitable tangible privacy designs can be leveraged to support a more comprehensive assessment of users' privacy in the context of smart voice assistants with embedded microphones.

In particular, we seek to answer the following research questions in this work:

**RQ1** How can tangible privacy mechanisms be integrated into the current design space of smart voice assistants in the context of embedded microphones?

**RQ2** Do tangible privacy mechanisms affect users' perceived reliability of, trust in, sense of risk, or sense of control over smart voice assistants?

**RQ3** Do tangible privacy mechanisms affect users' perceived usability of voice assistants? Is there a trade-off between usability and the metrics in RQ2?

We study these research questions in the context of six prototypical designs and derive important implications for the design of smart voice assistants to provide privacy in the context of embedded microphones. The rest of this paper is organized as follows. Section 2 discusses related work. Section 3 describes our methodology including the design of our user study. Sections 4 and 5 present our quantitative and qualitative findings respectively. Section 6 discusses the implications of our findings. Finally, Section 7 presents our conclusions.

## 2 RELATED WORK

In this section, we start by discussing prior research related to privacy concerns of smart voice assistants. Then we discuss the privacy preserving solutions currently used for conveying privacy to the users of these devices.

### 2.1 Privacy Concerns of Smart Voice Assistants

Prior works have investigated and identified users' privacy perceptions, norms, and concerns regarding the use of smart voice assistants [1, 2, 28, 29, 51, 59, 64, 94]. In their analysis of online reviews of smart voice assistants from U.S. based online shopping websites, Fruchter and Liccardi found reviewers expressed concerns about the amount, scope, and type of data collected by their devices [44]. Users were also worried that their sensitive conversations would be overheard by the devices. Lau et al. found that incidental users of smart voice assistant have an incomplete understanding of the resulting privacy risks [59]. They also reported that users preferred unplugging their devices instead of using the current privacy controls provided with the devices. Prior researchers have also studied users' perceptions of how smart voice assistants work and found a

[3] https://portal.facebook.com/help/479997345804912/

[4] https://support.apple.com/guide/security/hardware-microphone-disconnect-secbbd20b00b/web

significant knowledge gap between people's perception of these devices and how these devices actually work [5]. Abdi et al. showed that users have incomplete mental models of how personal voice assistants function, which often leads to different perceptions of where data is being stored, processed, and shared [1]. Similarly, Malkin et al. in their study found that almost half of their respondents were not aware of the use of cloud storage for their voice recordings, and even fewer had ever used the manufacturer provided privacy control of deleting their recordings from the cloud [64]. Participants in their study were also concerned about their recordings being used for advertising or being accessed by third-parties.

Tabassum et al. in their study found that participants in general are uncertain or unaware of the controls available on their devices, and such uncertainties led them to not use certain device functionalities; in extreme cases participants removed the device from their houses [95]. Often such lack of understanding of how these devices work not only acts as a barrier towards the adoption of these devices, but also contributes to users' privacy concerns [11, 13, 59]. Furthermore, Ammari et al. found that people have concerns about random false activation of smart voice assistants and third-party use of voice recordings [6]. Similarly, Huang et al. found that users perceive the collection of their voice recordings by external entities (e.g., smart speaker vendors) as a major privacy threat [51]. Participants also had privacy concerns such as the unauthorized access of their personal information and the misuse of their devices by other users such as visitors. Tabassum et al. found that people expressed concerns about the always-listening nature of smart voice assistants [94]. Chung et al. identified unintentional voice recordings as a large privacy threat stating that users are not provided with complete control over their voice data [31]. In their study, Zeng et al. found participants living in smart homes have privacy concerns related to audio recordings [105].

In contrast to these prior works, our work does not aim to further investigate users' privacy concerns. Rather, we aim to assess whether users think that hardware based tangible control and feedback mechanisms can successfully mitigate their privacy concerns and provide them with their desired privacy protection against unwanted audio recording.

## 2.2 Privacy Preserving Solutions for Smart Voice Assistants

To mitigate users' privacy concerns, researchers have proposed several system level solutions. Feng et al. proposed a continuous authentication system [39]. This system ensures the voice assistant executes the commands only of the device owner by matching the body surface vibrations of the user to the sound signal received by the voice assistant's microphone. Other researchers have focused on intercepting or monitoring network traffic in an attempt to address privacy and security issues [75, 91]. Although encryption of traffic can prevent such monitoring, prior research has shown that even encrypted traffic is vulnerable to privacy attacks [4, 7]. However, these approaches are orthogonal to our goals, which aim to provide 'high assurance' about the prevention of data collection itself. Moreover, Kumar et al. showed that network level solutions do not guard against skill squatting attacks [58]. Recently, smart voice assistant manufacturers have enabled users to access, review, and delete past voice recordings [34]. However, studies have shown that few users actually review or delete their past recordings, and many do not even know such option exists [6, 64]. Another stream of solution involves blocking or jamming the microphone of the voice assistants by using different forms of ultrasound [45, 62] or by continuously playing low frequency noise atop of the microphone of the smart speaker [54]. Chandrasekaran et al. proposed a similar privacy-preserving interventions, where the smart speaker's microphone is jammed using a separate remote control device [24]. Chen et al. also designed a jamming device that uses multiple ultrasonic transducers to disable microphones from all directions [27]. Olade et al. introduced an intermediary device that provides an additional layer of security by intelligently filtering sensitive conversations from being recorded by voice assistants [71]. Researchers have also suggested privacy controls based on interpersonal communication cues such as gaze direction and voice volume [66, 68]. However, jamming microphones using ultrasound does not provide users any tangible feedback that their privacy is protected and playing continuous noise in front of the speaker could be annoying for occupants near the smart voice assistants.

In summary, the current privacy preserving solutions provide some privacy protection to the users but most of them are software based solutions - hence, having trust issues and lacking effective feedback about whether privacy protection mechanisms are working properly (or not) [5], and people often find them cumbersome to use [59]. In contrast to these works, our work provides evidence for the adoption of tangible, on-device hardware based control and feedback mechanisms by assessing users' attitude towards tangible mechanisms.

## 3 METHOD: SURVEY STUDY

To answer our research questions, we conducted an online survey to understand several aspects of user attitudes towards smart voice assistants that incorporate tangible design principles. In the survey, we considered six different design prototypes of smart voice assistants by varying the type of control mechanism (physical control vs. software control) and feedback mechanism (physical feedback, vs. software feedback, or LED-based feedback). We designed the study as a between-subjects survey in which the participants were randomly assigned to one of the six possible conditions and were presented one of the six virtual prototypes of voice assistants. Participants took approximately 15—20 minutes to complete the survey, which was approved by our institution's ethics review board.

### 3.1 Experimental Conditions: Design Prototypes

Following a scenario-based design [82], we presented our participants with a hypothetical scenario. Our participants were asked to consider themselves as someone staying in a hotel room and having a private conversation with their friend in the visible presence of one of the six prototypical voice assistants. Below, we describe the design principles of the six prototypical voice assistants presented as virtual prototypes [21, 98] in this study. Our prototypes provide users with tangible (and non-tangible) ways of turning the embedded microphone on and off. They also provide tangible (and non-tangible) feedback about the microphone's state (whether it's listening or not).

***Physical Control & Physical Feedback.*** This prototype (Fig 1) provides users with the option to control the microphone using a physical hardware based on-off button. Unlike software based virtual buttons (as seen in mobile apps), sliding these physical buttons disables (and enables) the microphone, thus providing a tangible way of turning off (and on) the microphone. We deliberately used mute-switch based prototypes to study any differences between mute switches without assured feedback (existing voice assistants) and those with assured physical feedback (showing a physical disconnection). The physical feedback is provided by using a visible microphone jack. This jack is internally connected with the microphone. Whenever the microphone is turned off, the jack is visibly disconnected and whenever the microphone is on, the jack is visibly connected. Users can monitor this action of connection (and disconnection) of the microphone jack through a transparent cover.

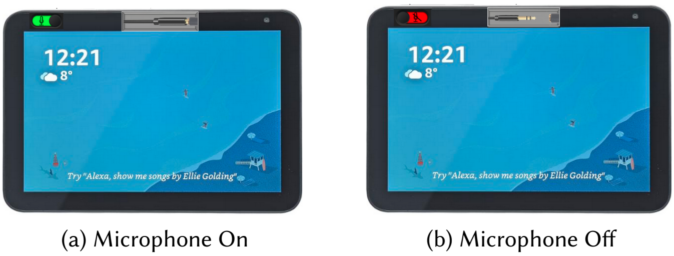


Fig. 1. Prototype with physical control and physical feedback

***Physical Control & LED Feedback.*** This prototype (Fig 2) provides a physical hardware based on-off button to control (turn on and off) the microphone. However, for feedback, this prototype uses an LED to communicate the current status of the microphone. For example, when the LED is flashing green, it means the microphone is 'on' and when the LED is flashing red, it indicates the microphone is 'off'.

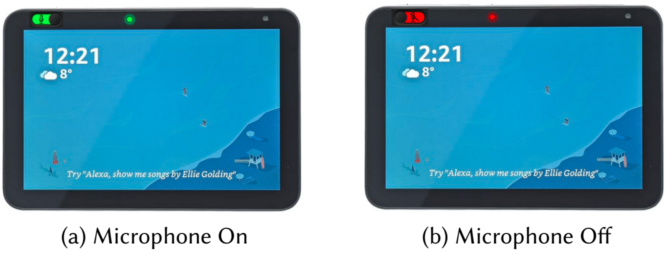
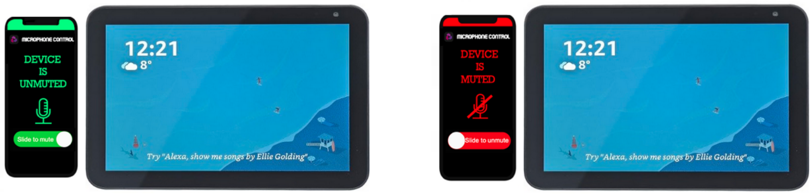


Fig. 2. Prototype with physical control and LED feedback

***Software Control & Software Feedback.*** This prototype (Fig 3) provides users with a combination of software based control and feedback mechanisms. The users need to turn the microphone on and off using the mobile app. They also need to rely on the app to notify them about the state of the microphone. The users see a textual description of the microphone's state. Specific colors (green for 'on' and red for 'off') are also used to help users easily understand the state. We chose to include a smartphone based app because of the prevalence of such apps for allowing users to control their IoT devices (e.g., the Nest App for camera and microphone controls[5]). Furthermore, a prior study by Chandrasekaran et al. showed that people found physical interactions with privacy preserving interventions for smart speakers to be less ideal and expressed their preference toward using an app on their smartphones for controlling the interventions [23].

[5] https://support.google.com/googlenest/answer/9210305



(a) Microphone On

(b) Microphone Off

Fig. 3. Prototype with software based control and feedback mechanism.

***Physical Control & Software Feedback.*** The prototype provides a physical hardware based on-off button to control (turn on and off) the microphone. The software based feedback is provided using a mobile app. Whenever the microphone's status changes, the user can see the notification on their mobile app in the form of a textual description of the microphone's current status. Different colors are used to indicate different statuses of the microphone. The choice of colors was grounded in a prior study by Ahmad et al. [5] – green was used to indicate the 'on' state, while red was used to indicate the 'off' state.

***Software Control & Physical Feedback.*** This design allows the microphone to be controlled only by using software based mechanisms. Users need to turn the microphone on and off by pushing a software based button in a mobile app. The on-off instruction is carried over the internet and the prototype's microphone changes its state accordingly. For providing physical feedback, just like the first prototype in figure 1, this prototype uses a microphone jack that comes with a transparent cover. This jack is connected and disconnected based on the instruction it receives from the mobile app of the user.

***Software Control & LED Feedback.*** The last prototype provides users with software based controlling mechanisms along with LED based feedback. Just like the previous combinations, the software based control involves pushing a virtual button in a mobile app and the LED changes its color according to the microphone's current state. The only notification the user will get is from the LED – flashing green when the microphone is 'on' and flashing red when it's 'off'.

### 3.2 Measurements: Independent and Dependent Variables

We used two independent variables in our study as indicated above. One is the 'control mechanism' with two levels (hardware based/physical and software based). The other is the 'feedback mechanism' with three levels (hardware based/physical, software based, and LED based). Table 1 gives an overview of our experimental conditions. We measured five dependent variables for each condition: perceived usability, perceived reliability, perceived trust, perceived control, and perceived risk. We measured these variables because prior works have shown these constructs influence users' privacy perceptions [14, 88] and their willingness of smart device adoption [22, 38, 89]. Next, we describe our dependent variables.

| **Independent Variables** | **Dependent Variables** |
|---|---|
| Control mechanism (Physical, Software) | Reliability, Trust, Control, |
| Feedback mechanism (Physical, Software, LED) | Risk Perception, Usability |

Table 1. Visual representation of the variables used in our experimental design. The first column shows our independent variables with their different levels. The second column shows the dependent variables that we measured for each of the experimental conditions.

*3.2.1 Perceived Usability.* We asked the participants to rate their perceived usability of the prototypes by using a 10 item questionnaire from the System Usability Scale (SUS) developed by Brooke et al [15]. The items were modified to fit the context of our study. Example items include: "I found this approach to mute the microphone unnecessarily complex". The 10 items are scored on a 5-point Likert scale of strength of agreement (1: strongly disagree; 5: strongly agree).

*3.2.2 Perceived Reliability.* Three items were borrowed and adapted from Madsen and Gregor's [63] perceived reliability scale to fit the context of measuring the reliability of muting process of smart voice assistants. A sample item is "Using this approach to mute the microphone, I can rely on the device to not record me". We used a 5-point Likert scale ranging from (1: strongly disagree; 5: strongly agree) to collect participants' level of agreement with the three adapted scale items.

*3.2.3 Perceived Trust.* To measure participants' perceived trust on the prototypes, we borrowed eight items from the trust in automation scale developed by Jian et al [52]. We changed the wordings of the scale items to fit the context of our study. Example items include: "I am skeptical of the visual information provided by this approach of muting the microphone", "I can trust this approach of muting the microphone for preventing unwanted audio recording". Participants were asked to rate their level of agreement with the eight statements using a 5-point Likert scale ranging from (1: strongly disagree; 5: strongly agree).

*3.2.4 Perceived Control.* To measure participants' perceived control on the prototypes, we modified the two-item scale developed by Hinds [47] to fit the context of how much the participants feel in control after muting microphone. A sample item is "With this approach of muting the microphone, I feel I am in control of any recordings by this device". Participants indicated their level of perceived control on the prototypes on a 5-point Likert scale ranging from (1: strongly disagree; 5: strongly agree).

*3.2.5 Risk Perception.* We measured participants' risk perception of interacting with the prototypes by adapting a four-item scale from the work of Gulati et al. [46] and Colesca [33]. Items that best fit the assessment of participants' risk perception after muting a smart voice assistant were chosen and modified to fit the study purpose. An example item is "When I mute the microphone using this approach, I feel it would be risky to have a private conversation around this device". Participants were asked to rate the four adapted items using a 5-point Likert scale ranging from (1: strongly disagree; 5: strongly agree).

### 3.3 Measurements: Covariates

*3.3.1 Technical competence.* Technical competence is defined as the confidence in one's own ability to solve technical problems. Prior studies found technical competence to be an important technology-related personality trait that has a strong association with technology acceptance [12, 83]. We measured participants' technical competence using the shorter version of the subjective technical competence scale used by Arning et al [9]. This scale has eight items and these items had to be confirmed or denied on a 5-point Likert scale from 1 (totally disagree) to 5 (totally agree).

*3.3.2 Privacy Self-efficacy, Privacy Awareness, and Privacy Preference.* Participants' privacy self-efficacy was measured using a six-item scale ($\alpha = 0.691$). Their privacy awareness was measured using a three-item scale ($\alpha = 0.637$). Both of these scales were adapted from Zeissig et al [104] and the items were rated on a 5-point Likert scale from 1 (totally disagree) to 5 (totally agree). Privacy self-efficacy has been shown to influence privacy protection behaviors [26] and privacy awareness influences privacy concerns [101]. Finally, a one item privacy preference question was used, which asked participants, "Are you a private person who keeps to yourself or an open person who enjoys sharing with others? (1 = very private, 7 = very open)" [50]. We conducted an Exploratory Factor Analysis (EFA) on these three privacy metrics to discover the underlying concepts (Section 3.7.2) and came up with four concepts: privacy awareness, privacy confidence, privacy vigilance, and privacy comprehension. We used these four concepts as the covariates in our analysis.

### 3.4 Survey design

Our survey was comprised of both open-ended and close-ended questions and was approved by our institution's ethics review board. We organized the survey instrument as follows: (Our survey questionnaires are presented in Appendix A.)

- Consent form.
- Questions about participants' experience of using smart voice assistants, e.g., which voice assistants participants have used and how often they use such devices.
- Presentation of the survey scenario and one of six design prototypes based on random assignment (between-subjects).
- A total of 27 questions to understand participants' perceived usability, reliability, trust, control, and risk about the assigned prototype.
- Open ended questions about why participants found the prototype to be reliable (or not), easy to use (or not), and the features of the prototype they liked (or not).
- Questions for measuring the participant's technical competence and privacy information such as privacy efficacy, privacy awareness, and privacy preference.
- Four demographic questions (age, gender identity, race/ethnicity identity, and education).

### 3.5 Recruitment, Screening, and Compensation

The survey was implemented in Qualtrics [80] and advertised on Amazon Mechanical Turk (MTurk) [17]. We hosted the survey for over a period of two months. It was restricted to MTurk workers who were at least 18 years old, native English speakers, had experience using smart voice assistants, and had been living in the USA for at least five years (to help control for cultural variability [55]). We further required that workers have a high reputation (approval rate of greater than or equal to 99% on at least 1,000 completed HITs) to ensure data quality [67]. We also added a captcha at the beginning of the survey and used three attention-check questions to filter out inattentive responses [61]. Each participant could participate in the survey only once. In our data analysis we included the responses of 261 participants (out of a total of 270) who correctly answered all three attention-check questions and entered the correct random response code as generated by the survey instrument. We discarded the responses of nine participants who failed at least one out of the three attention check questions. Each participant was paid $3.00, whether or not we used their responses. The amount was determined through a pilot study where participants were asked whether they considered the compensation to be fair. Participants were able to pause this survey and resume at a later time, as indicated by the long completion time (>10hours) for a number of participants. Therefore we analyzed the response times for the top 75%, which was an average of 16 minutes. Thus we estimated that our compensation was in the range of around $11/hour for the work in our survey. The study and compensation scheme was approved by our institution's ethics review board.

### 3.6 Pilot Study

We conducted in-person online surveys and follow-up interviews with two male participants to identify any accessibility issues of our survey instrument. Their ages ranged from 20 to 30. Both participants participated in the survey using computers. The pilot study took around 30—40 minutes for each participant. Participants were compensated with $20 cash for taking part in the pilot. We requested them to point out any accessibility issues they faced while participating in the survey. We also requested that they suggest improvements to our survey. During the in-person online survey with the first participant, we identified varying levels of accessibility issues such as difficulties in navigating through the text fields, a number of typos, and minor confusion about the wording of some questions. We addressed these issues mentioned by the first participant and conducted the second in-person online survey a few days later. The second participant did not raise any accessibility issues and thus we finalized the survey. During both the follow-up interviews, our goal was to find out whether the participants could understand the description of the prototypes and the hypothetical scenario they were put in. We also asked for their suggestions to make the wording easier to understand. Both the participants suggested minor modifications and we modified the prototype description based on their suggestions. Apart from this, we also conducted a pilot study with 5 participants on Mturk to make sure that the survey ran smoothly.

## 3.7 Analysis Procedures

We now describe our quantitative and qualitative analysis procedures.

*3.7.1* ***Quantitative analysis.*** For our statistical quantitative analysis, we used multiple linear regression to model the relationship between our dependent and independent variables. We designed the models to allow us to explore how the independent variables, e.g. control mechanism and feedback mechanism, influenced our dependent variables. Along with the independent variables, the linear models that we created had several combinations of the co-variates (Section 3.3). We selected a subset of the co-variates that represented the best fitting model based on the AIC values [84]. For measuring the dependent variables, we calculated the factor loadings (Table 6) of each item in the scales and used them in our analysis. Our data show that the dependent variables are not normally distributed. Hence, we chose to use a simple linear regression because (a) it does not require the normality assumption to be fulfilled and (b) it is a sound statistical technique in determining the degree to which particular independent variables are influencing dependent variables.

Apart from this, we also conducted exploratory factor analysis (EFA) on the privacy metrics to extract any underlying privacy concepts. EFA helped to group the correlated features under a common factor based on the absolute values of factor loadings. Informed by the multiple linear regression and factor analyses, we identified multiple subsets of features that show how the dependent variables are affected by the independent variables. In addition to a simple linear regression, we also used logistic regressions in our analysis. Since we had similar outcomes for both the linear and logistic regressions, we chose to report the results of linear regression for ease of interpretation.

*3.7.2 Uncovering Underlying Factors of Privacy Metrics.* During the survey we collected data for measuring three privacy related scales: privacy self-efficacy, privacy awareness, and privacy preference. As the three metrics are closely related to each other, we wanted to ensure that the subset of selected features are minimally correlated with each other since multicollinearity can result in an unstable linear model [40]. So, we conducted an exploratory factor analysis (EFA) on these three privacy metrics to discover the underlying concepts that represent a set of variables that are minimally correlated among themselves and retain maximum variance of the outcome variable. We identified four underlying variables and these variables were calculated using the factor loadings shown in Table 5 in Appendix A.

- Determining eligibility of EFA: We used the Kaiser-Meyer-Olkin (KMO) value to determine whether our data is suited for factor analysis. The total KMO was 0.77, which based on

Kaiser's suggested cutoff (KMO $\geq 0.60$) indicates that, we can conduct a factor analysis. We also conducted Bartlett's Test of Sphericity ($p = 2.105859e - 180$), which indicated that a factor analysis may be useful with our data.

- Determining the number of factors to extract: We conducted principal component analysis (PCA) to estimate the amount of variance retained by each component. We decided the number of factors to extract from EFA using a scree plot [72, 103].
- Extracting and rotating factors: After deciding on the number of factors, we extracted the factors and estimated the factor loading (i.e., correlation between a feature and a factor) of each feature. Finally, we rotated the factors using 'varimax' rotation to obtain a simple structure of the factor loadings [72, 103]. It helps to group the features and since ideally each feature has a high factor loading for only one factor after the rotation, it also helps to clearly describe the features. Features that are highly correlated among themselves belong to the same underlying concept and would have high correlation with that concept. Consequently, we grouped the features having high correlation with a single factor into categories describing 'meaningful' constructs. Thus, we identified four underlying concepts: privacy comprehension, privacy confidence, privacy vigilance, and privacy awareness, which we use as the covariates.

*3.7.3 Sample size power analysis.* We performed a power analysis to estimate the sample size required to produce statistically significant findings for each dependent variable. The analysis showed that for testing the effect of the two independent variables on each dependent variable, 107 participants would provide enough statistical power to detect 0.15 ('small') sized effects ($\alpha = 0.05$, 1-$\beta = 0.95$).

*3.7.4* ***Qualitative analysis.*** Participants' responses were recorded on three open-ended questions related to their perceived ease of use, reliability, useful features of the prototypes, and their line of reasoning for the responses. All qualitative answers were independently coded in a bottom-up approach by two researchers. Regular discussions were held between the two researchers about the codes and themes throughout the qualitative analysis. Initially, the researchers met weekly to iteratively and redundantly code a subset of the open-ended responses from the survey to develop a code-book. Then, based on the code-book, we coded the rest of the responses. We chose to perform a thematic analysis because although our work is informed by relevant theories such as user-privacy theories, (a) we did not directly derive our coding from those theories, and (b) we did not set our goal to build new theories in our work. We came up with three main themes which are described in Section 5. In line with common practice, we did not seek to compute inter-rater reliability since we focused on a thematic analysis based on multiple iterations of meetings and refinement of the codes to determine emergent themes (these codes were not used in our quantitative analysis) [8, 65].

## 3.8 Hypotheses

In this section, we present the two hypotheses that we seek to test in this study. Each of the two hypotheses has three sub-hypotheses.

- H1 (a/b/c): Hardware based control mechanisms have higher perceived [usability/control/trust] than software based control mechanisms.
- H2 (a/b/c): Hardware based feedback mechanisms have higher perceived [usability/control/trust] than software based feedback mechanisms.

Due to high correlation between perceived reliability, risk perception, and perceived trust, we decided to test the hypotheses on three dependent variables (explained in 4.1).

# 4 QUANTITATIVE FINDINGS

In this section, we present the findings of our statistical analyses.

## 4.1 Correlation Between Perceived Reliability, Risk Perception, and Perceived Trust

We found a high correlation among the three dependent variables perceived reliability, risk perception, and perceived trust. Spearman's correlation tests showed high correlations between perceived reliability and perceived trust ($rho = 0.8896246, S = 327067, p < 2.2e - 16$), and high negative correlations between perceived trust and risk perception ($rho = -0.9030382, S = 5639121, p < 2.2e - 16$) and perceived reliability and risk perception ($rho = -0.7866527, S = 5294245, p < 2.2e - 16$). Therefore, we focus only on perceived trust in the presentation of our analysis. The related analyses for perceived reliability and risk perception are included in table 7 in Appendix A.

## 4.2 Overall Effects of Control and Feedback Mechanisms

In this section we describe the results of our hypothesis testing i.e., how the independent variables along with the co-variates effected the dependent variables.

*4.2.1* ***Factors Affecting Perceived Usability.*** To test H1(a) and H2(a) we use a linear regression model containing two independent variables (control and feedback mechanism) and three co-variates (technical competence, privacy awareness, and privacy comprehension). The three co-variates were chosen based on Akaike's Information Criterion (AIC) [84], i.e., the model with these three co-variates had the lowest AIC value and this model best fit our data. The model shows there was a significant effect of control mechanism on perceived usability ($p = 1.55e - 11$) with a 10.738 unit increase (on a scale of 1–100) in perceived usability when the control mechanism was changed from software based to hardware based (Table 2). In other words, in line with the findings of prior works [41, 56], our participants also found hardware based tangible control mechanisms to be more usable compared to software based ones and thus, we accept H1(a).

However, we do not accept H2(a) as there was no significant effect of feedback mechanisms on perceived usability. Among the co-variates, privacy comprehension is significantly associated with perceived usability ($p = 7.88e - 05$). The model indicates a 9.881 unit decrease in perceived usability for every unit increase in privacy comprehension. Figure 4 shows the overall distribution of participants' ratings of perceived usability of the six combinations of control and feedback mechanism. The key insight highlighted in this figure is that the usability rating for all combinations with physical control is considerably higher compared to combinations with software control (further discussed in Section 6). Interestingly, not a single participant disagreed that the combination of physical control and LED feedback increase usability. This might be because recent smart voice assistants come with push buttons and LEDs, and participants rating for usability might be influenced by being habituated to seeing these devices.

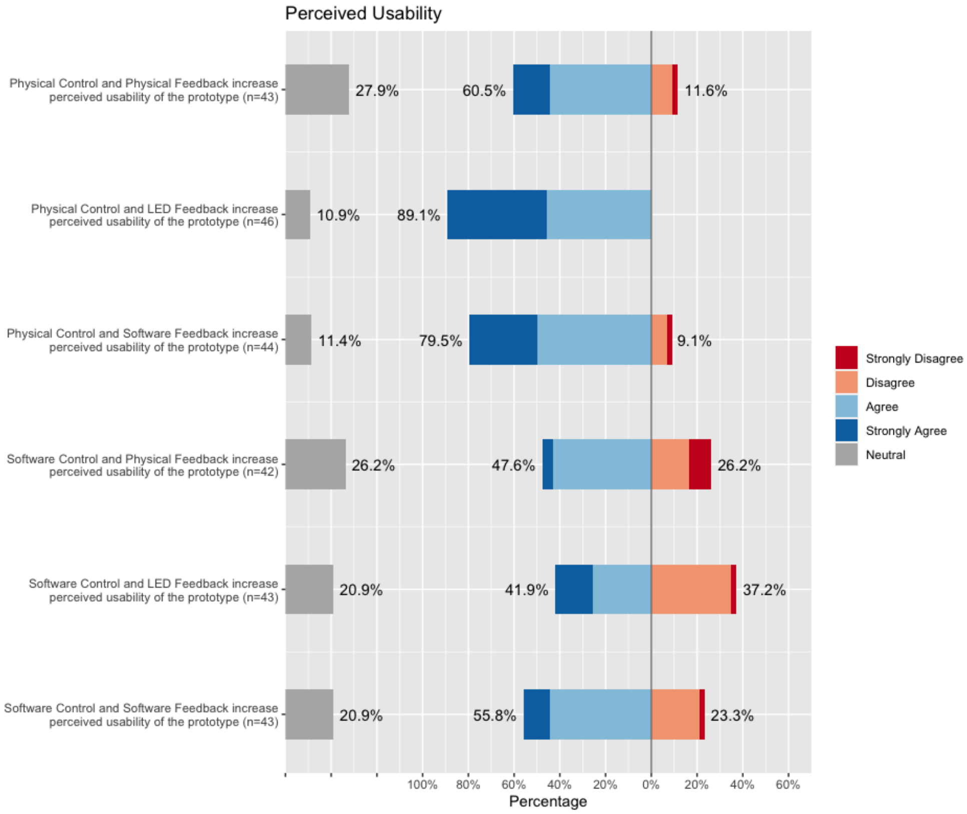


Fig. 4. Distribution of participants' rating for Perceived Usability across six combinations of control and feedback mechanism

*4.2.2* ***Factors Affecting Perceived Control***. For testing H1(b) and H2(b) we used a linear regression model containing the two independent variables control and feedback mechanism and three co-variates privacy confidence, privacy vigilance, and privacy comprehension. This model with these three co-variates was chosen based on the lowest Akaike's Information Criterion (AIC) value. Our result indicate there was a significant effect of control mechanism over perceived control ($p = 1.37e - 05$) – the model indicates a 0.513 unit increase in perceived control when the control mechanism was changed from software based mechanisms to hardware based mechanisms (Table 2). Thus we accept H1(b) as participants' perceived increased control over the device with hardware based control mechanisms compared to software based control mechanisms.

However, H2(b) is not accepted as we found no significant relation between feedback mechanism and perceived control. All three co-variates, privacy confidence ($p = 0.00177$), privacy comprehension ($p = 0.00462$), and privacy vigilance ($p = 0.03191$) were significantly associated with perceived control. The model indicates a decrease of 0.143 and 0.432 units in perceived control for every unit increase in privacy vigilance and privacy comprehension respectively. Interestingly, there was a 0.516 unit increase in perceived control for every unit increase in privacy confidence.

*4.2.3* ***Factors Affecting Perceived Trust***. For testing H1(c) and H2(c) we used a linear regression model containing the two independent variables control and feedback mechanism and the three co-variates privacy confidence, privacy vigilance, and privacy comprehension. Just like the previous models, this model was chosen based on the lowest Akaike's Information Criterion (AIC) value. Control mechanism had a significant effect over perceived trust ($p = 5.68e - 05$) – the model indicates a 0.413 unit increase in perceived trust when the control mechanism was changed from software based mechanisms to hardware based mechanisms (Table 2). Thus H1(c) is accepted as hardware based control mechanisms increase participants' perceived trust on the device compared to software based control mechanisms.

H2(c) is not accepted as we found no significant relation between feedback mechanism and perceived trust. All three co-variates, privacy confidence ($p = 2.84e - 06$), privacy comprehension ($p = 0.00918$), and privacy vigilance ($p = 2.84e - 05$) were significantly associated with perceived control. The model indicates a decrease of 0.247 and 0.346 units in perceived trust for every unit increase in privacy vigilance and privacy comprehension respectively. Interestingly, there was a 0.682 unit increase in perceived trust for each unit increase in privacy confidence.

### 4.3 Interaction Effects of Control and Feedback Mechanisms

We wanted to assess if there was any significant interaction between the control mechanism and feedback mechanism for the dependent variables. We used simple linear models with interaction terms between control and feedback mechanism. As shown in Table 3, there was no significant interaction between control and feedback mechanism for perceived control and perceived trust. However, in terms of usability, there was a significant interaction between physical control mechanism and LED feedback. The results indicate that switching to physical control (from software control) with LED feedback results in higher perceived usability than switching to physical control with software feedback. In other words, participants found the combination of physical control with LED feedback mechanism to be more usable compared to the other combinations. Overall we can say that just like the main effect, the interaction of feedback mechanism and control mechanism does not have any significant effect on perceived control and trust. However, LED feedback when paired with physical control mechanisms have significant interaction effect on usability.

### 4.4 Demographics

Our final sample of 261 participants were slightly skewed towards younger generation (under the age of 30), most of them identified themselves as males, and most of them used voice assistants several times a day. A brief description of the demographics is presented in Table 4.

### 4.5 Additional Factors

We wanted to assess the effect of gender and age on our dependent variables as these factors are found to be important in the context of privacy [25, 86, 87]. Past works have demonstrated that compared to men, women are more risk averse [25] and sensitive to privacy concerns [86]. Also, it was found that individuals older than 45 respond differently to privacy concerns than younger individuals [87]. We conducted a Wilcoxon rank sum test (between subject, two groups) for all the dependent variables and found no significant difference in the ratings based on gender. To simplify the age analysis, we categorized the participants into three age groups: 18–29, 30-49, and 50 and older. We conducted a Kruskal Wallis test (between subject, three groups) for the five dependent variables and the results show that the difference between the three age groups is not statistically significant for any of the dependent variables.

[6] Both Amazon Echo devices and Google Home

| | *Dependent variable:* | | |
|---|---|---|---|
| | Perceived Control | Perceived Usability | Perceived Trust |
| | (1) | (2) | (3) |
| privacy confidence | $0.516^{***}$ | | $0.682^{***}$ |
| | (0.196,0.836) | | (0.403,0.961) |
| | $p = 0.002$ | | $p = 0.00001$ |
| privacy vigilance | $-0.143^{**}$ | | $-0.247^{***}$ |
| | (-0.273,-0.013) | | (-0.360,-0.133) |
| | $p = 0.032$ | | $p = 0.00003$ |
| technical competence | | 2.591 | |
| | | (-0.271,5.453) | |
| | | $p = 0.078$ | |
| privacy awareness | | $-1.616$ | |
| | | (-3.745,0.514) | |
| | | $p = 0.139$ | |
| privacy comprehension | $-0.432^{***}$ | $-9.881^{***}$ | $-0.346^{***}$ |
| | (-0.729,-0.136) | (-14.706,-5.056) | (-0.605,-0.088) |
| | $p = 0.005$ | $p = 0.0001$ | $p = 0.010$ |
| physical control | $0.513^{***}$ | $10.738^{***}$ | $0.413^{***}$ |
| | (0.286,0.740) | (7.759,13.718) | (0.215,0.611) |
| | $p = 0.00002$ | $p = 0.000$ | $p = 0.0001$ |
| physical feedback | 0.083 | $-1.841$ | 0.109 |
| | (-0.199,0.366) | (-5.565,1.883) | (-0.137,0.356) |
| | $p = 0.564$ | $p = 0.334$ | $p = 0.385$ |
| LED feedback | $-0.014$ | 0.466 | 0.023 |
| | (-0.287,0.259) | (-3.130,4.062) | (-0.215,0.261) |
| | $p = 0.921$ | $p = 0.800$ | $p = 0.848$ |
| Constant | $3.473^{***}$ | $65.598^{***}$ | $2.702^{***}$ |
| | (2.694,4.252) | (48.257,82.939) | (2.023,3.381) |
| | $p = 0.000$ | $p = 0.000$ | $p = 0.000$ |
| Observations | 261 | 261 | 261 |
| R$^2$ | 0.144 | 0.274 | 0.202 |
| Adjusted R$^2$ | 0.124 | 0.257 | 0.183 |
| Residual Std. Error (df = 254) | 0.922 | 12.164 | 0.804 |
| F Statistic (df = 6; 254) | $7.148^{***}$ | $15.986^{***}$ | $10.699^{***}$ |
| *Note:* | | | $^{**}$p<0.05; $^{***}$p<0.01 |

Table 2. Regression Results (main effect) for perceived control, perceived usability, and perceived trust. Control mechanism has significant effect on perceived control ($p = 1.37e - 05$), perceived usability ($p = 1.55e - 11$), and perceived trust ($p = 5.68e - 05$).

## 5 QUALITATIVE FINDINGS

To enrich our understanding of what motivated the ratings provided by our participants, we coded the open-ended responses for additional insights. These responses helped us better understand

| | *Dependent variable:* | | |
|---|---|---|---|
| | Perceived Control | Perceived Usability | Perceived Trust |
| | (1) | (2) | (3) |
| physical control | $0.527^{**}$ | $7.605^{***}$ | $0.433^{**}$ |
| | (0.125,0.928) | (2.160,13.051) | (0.070,0.796) |
| | $p = 0.011$ | $p = 0.007$ | $p = 0.021$ |
| physical feedback | $-0.087$ | $-3.950$ | 0.040 |
| | (-0.494,0.319) | (-9.459,1.560) | (-0.327,0.408) |
| | $p = 0.674$ | $p = 0.162$ | $p = 0.831$ |
| LED feedback | 0.047 | $-4.016$ | $-0.031$ |
| | (-0.357,0.450) | (-9.493,1.461) | (-0.396,0.335) |
| | $p = 0.822$ | $p = 0.152$ | $p = 0.870$ |
| physical control:physical feedback | 0.131 | $-0.960$ | $-0.039$ |
| | (-0.441,0.702) | (-8.707,6.786) | (-0.556,0.477) |
| | $p = 0.655$ | $p = 0.809$ | $p = 0.882$ |
| physical control:LED feedback | $-0.084$ | $8.204^{**}$ | 0.161 |
| | (-0.648,0.481) | (0.544,15.864) | (-0.350,0.672) |
| | $p = 0.773$ | $p = 0.037$ | $p = 0.538$ |
| Constant | $3.326^{***}$ | $62.913^{***}$ | $2.539^{***}$ |
| | (3.040,3.611) | (59.040,66.785) | (2.281,2.798) |
| | $p = 0.000$ | $p = 0.000$ | $p = 0.000$ |
| Observations | 261 | 261 | 261 |
| $R^2$ | 0.078 | 0.173 | 0.075 |
| Adjusted $R^2$ | 0.060 | 0.157 | 0.057 |
| Residual Std. Error (df = 255) | 0.956 | 12.957 | 0.864 |
| F Statistic (df = 5; 255) | $4.297^{***}$ | $10.678^{***}$ | $4.115^{***}$ |
| *Note:* | | | $^{**}$p<0.05; $^{***}$p<0.01 |

Table 3. Regression Results (interaction effect) for perceived control, perceived usability, and perceived trust. The interaction between physical control and LED feedback have significant effect ($p = 0.037$) on perceived usability.

why our participants preferred hardware based mechanisms. This qualitative analysis also provides key design insights on how to incorporate tangible mechanisms into smart voice assistants. Below, we present the main themes identified in our qualitative analysis.

### 5.1 Hardware based mechanisms are reliable, coherent and easy to use

It was evident from the open-ended responses of our participants that hardware based control and feedback mechanisms are more reliable, coherent, easy to use, and provide them with a true sense of security against unwanted audio recording. For each of our examined construct, we provide sample quotes from our data to demonstrate the supporting evidence. Out of all the participants who were presented with hardware based control mechanisms, 68% of them mentioned hardware based control mechanisms as reliable in their open-ended feedback

***Reliability***. Comments from our participants highlighted why they found tangible control mechanisms to be more reliable. The physicality of a hardware switch and its mechanical way of controlling audio recording were described as the main reasons for higher perceived reliability of tangible control mechanisms by our participants.

> "I believe that this approach is a reliable way of prevention unwanted audio because the user has to physically do it." (P16)

> "It is very reliable because it disables the hardware in a mechanical way." (P246)

> "I believe it is reliable. Hardware is always more reliable than software." (P174)

***A sense of security***. The responses of our participants highlight that prototypes equipped with physical control mechanisms provided them with a true sense of security in terms of their privacy management. Unlike software control mechanisms, physically interacting with a tangible control mechanism like a hardware switch gave them this sense of security as claimed by P147.

> "I can see a feeling of security being given to the user by them having to physically interact with the assistant, so they're not stuck wondering if their mute command worked or not." (P147)

> "It's relatively easy to do and provides a little bit of a sense of security." (P162)

***Ease of use***. From a usability perspective, participants mentioned the use of a physical switch is easy to understand. They also mentioned that muting a microphone using a physical switch saves a lot of time compared to going through a number of steps in a software application.

> "I found this approach of muting the microphone very easy to use as the process is intuitive." (P192)

> "The fact that it was done physically made it even easier than navigating a menu or telling the device to mute." (P88)

In summary, the responses from our participants provide clear evidence that for muting smart voice assistants, they preferred tangible control mechanisms such as a physical switch.

| **Gender** | | **Age** | | **Race** | |
|---|---|---|---|---|---|
| Female | 101 (38.7%) | 18-29 | 53 (20.3%) | White | 205 (78.54%) |
| Male | 157 (60.1%) | 30-49 | 165 (63.22%) | African-American | 20 (7.66%) |
| Non-binary | 3 (1.1%) | 50-64 | 36 (13.8%) | Asian | 15 (5.74%) |
| | | >65 | 7 (2.68%) | Hispanic | 6 (2.29%) |
| | | | | Others | 15 (5.74%) |

| **Education** | | **Device Type** | | **Device Usage** | |
|---|---|---|---|---|---|
| PhD | 5 (1.92%) | Amazon Echo | 105 (40.22%) | Several times a day | 111 (42.5%) |
| Bachelors | 136 (52.1%) | Google Home | 69 (26.43%) | About once a day | 40 (15.33%) |
| Masters | 33 (12.64%) | Both[6] | 69 (26.43%) | Few times a week | 63 (24.13%) |
| Diploma | 76 (29.12%) | Others | 18 (6.89%) | Few times a month | 29 (11.11%) |
| Trade-School | 7 (2.68%) | | | Few times a year | 18 (6.89%) |
| Other | 4 (1.53%) | | | | |

Table 4. Visual representation of participant demographics.

### 5.2 Software based mechanisms are cumbersome, deceptive, and untrustworthy

In general our participants expressed their distrust towards software based control and feedback mechanisms. For example, P70 was suspicious about software based feedback mechanisms:

> "I only trust a hardware switch. I can't tell [if] this is true [reliable] just because software says it's so." (P70)

The open-ended responses from our participants provided additional insight to why software-based feedback was not perceived as positively by them. Below, we highlight the main themes emerging in those comments:

***Providing a false sense of privacy.*** In contrast to tangible control and feedback mechanisms, participants were concerned that software based control mechanisms could fail and hence give them a false sense of privacy.

> "I feel that the app could fail and give a false sense of privacy." (P135)

It was evident from their responses that they found software based controls not trustworthy. Out of all the participants who were presented with a software based control mechanism, 48% of them mentioned software based control mechanisms as untrustworthy in their open-ended feedback.

> "Even though it says it is muted, I would always wonder if it is recording as long as it's on." (P219)

Participants said that software based mechanisms may malfunction without any notice and they preferred hardware based mechanical switches to software switches which align with the findings of prior work [5].

> "The technology may fail without any notice, would need to test each time to actually trust it and then it still may malfunction." (P141)

> "I wouldn't trust this. I prefer mechanical switches for these rather than soft[ware] switches." (P148)

***LEDs can be deceptive.*** Interestingly, although LED based feedback mechanisms were rated to be more usable by our participants, many of them were aware that LEDs are not reliable indicators and reported that they won't believe the feedback provided by LEDs.

> "I am aware that status lights are not reliable indicators of whether a device is active or not." (P117)

> "I thought the light showing the microphone was off was useful, but I'm not sure I can believe it is actually turned off." (P20)

***Usability issues.*** We identified several usability issues with software based control mechanisms from the responses of our participants. For software application based control mechanisms, users need to navigate through multiple steps using their mobile phone to mute the microphone. Participants found this process to be cumbersome and complicated as mentioned by P113 and P158.

> "Honestly, I think it's just way too cumbersome. If you really want to mute a device, unplug it." (P158)

> "I thought it was too complicated; there should be an option on the device itself."(P113)

Other participants mentioned general usability issues regarding the use of software applications from another device such as a mobile phone. For example, P107 said that people would be confused about the muting process of the microphone using a software application.

> "I think it will be too confusing for most people. It should not require your mobile device." (P107)

Overall, participants were skeptical about the effectiveness of using software based mechanisms to prevent unwanted audio recording. They also had similar concerns about using LEDs for feedback.

### 5.3 Desire for on-device, transparent control and feedback mechanisms

The open-ended responses of our participants revealed several sought out properties of control and feedback mechanisms in smart devices. Next, we present some of the important properties identified during the analysis.

***On-device and easy to access mechanisms.*** We found that participants preferred to have on-device control mechanisms. They also wanted to have feedback about the device's state from the device itself rather than depending on some other mechanism. Participants doubted the reliability of controlling the microphone using software applications and clearly asked for on-device controlling mechanisms.

> "I would only consider muting the device directly from the device to be reliable."(P144)

> "I don't like that I have to use my phone, it makes me feel like this device is able to sync with my phone and invade my privacy. I want to be able to mute it just using the device." (P105)

Other participants were not sure about the reliability of the prototypes as the feedback about the devices' states were not provided from within the devices.

> "I am not sure [of its reliability], because the device itself does not show that it is muted. If it did I might believe it for sure." (P253)

> "No [not reliable], I need confirmation from the actual listening device." (P43)

Another issue was the ease of accessing controlling mechanisms. Instead of obscured mechanisms, they wanted controlling mechanisms that are easy to find. They rated prototypes with easily accessible controlling mechanisms to be more reliable.

> "Rather than hide the mute option behind layers and steps of menus, by putting the slider button right at the top right of the screen, it is easy to access, and you could tell someone else how to do it in one step." (P176)

> "I believe it's reliable, they have it there that way it's easier to find and mute it that way instead of unplugging the device." (P120)

***Transparent and self explanatory.*** Transparency was a commonly sought out property among our participants. While talking about the hardware based feedback mechanism, they mentioned seeing the microphone get physically disconnected gave them peace of mind.

> "The clear plastic window showing the microphone being physically disconnected brings a little more peace of mind." (P87)

> "I think it is very reliable because it is transparent. I can see for myself that the mic is definitely not connected." (P29)

Along with transparency, participants also liked the prototypes that had easy to use control mechanisms and self-explanatory feedback mechanisms.

> "[I like it because] It is very self-explanatory and the graphics displayed on the button are very helpful." (P29)

> "I think it's reliable as it gives you an easier way to control what the mic hears or doesn't hear." (P63)

***I am the master of my device.*** One interesting finding was participants' wish to exert physical control over their devices by themselves. They liked the use of physical push button as it allows them to take the device's control into their own hand. They rated the prototypes with physical control mechanism as reliable because they could mute the microphone on their own instead of using or relying on any software application.

> "Rather than relying on the machine to automatically mute the microphone, I can do it myself with the push of a button." (P4)

> "I think this is reliable because I am physically disabling the device/microphone and not relying on the software/AI to mute the microphone." (P68)

In summary, our findings point to participants' desire of having devices that provide feedback and control mechanisms which are clear to understand and easy to manipulate, thereby giving them a sense of agency over their devices.

## 6 DISCUSSION

In this section we discuss a) the effectiveness of tangible control and feedback mechanisms on usability and user's sense of agency, b) the consistency of our results with paradoxical behavior noted in past literature, c) the relationship between usability and privacy, and d) limitations of our study.

### 6.1 Tangible Control vs. Feedback

An important theme identified in our qualitative analysis is that users want to be in total control of their devices. This theme is related to the idea of users' 'sense of agency', which is defined as the experience of being in control of one's own device [60]. In general, it has been shown that users strongly desire the sense that they are in charge of their device, and their interactions with devices also increase when they feel superior to their device [90]. Limerick et al. found that hardware based interaction increases users' sense of agency compared to other modalities of interaction [60]. Based on our findings that hardware control mechanisms improve usability and perceived trust, we recommend that future designs of smart voice assistants should indeed incorporate tangible control mechanisms to provide users with an increased sense of agency.

Interestingly, our findings show that the feedback mechanism did not have any significant effect on perceived reliability, trust, or risk perception of the device. This is an interesting finding considering that prior studies have shown that people expect visual feedback [19, 30] from digital devices and according to Norman [69] feedback about the internal operations of systems is essential for raising awareness, reassurance, and anticipation of further actions among the users. To get some perspective regarding this finding, we looked at our open-ended responses and interestingly, found that whenever participants mentioned different types of feedback mechanisms, they were not thinking about the impact of feedback mechanism on the reliability of the device. Rather, they discussed its impact on usability of the device (e.g., whether it was easy to interpret, placed in a convenient location, and so on). For example, regarding the LED based feedback mechanism, P47 said, "The LED makes it easier from a glance or farther away to tell if it's muted or not". The reason could be in the survey scenario, participants were allowed to use the control mechanisms of the prototype and thus focused on the control mechanisms in terms of measuring the reliability of the prototype. Although, device owners may not need feedback based on the control mechanisms used, it may be difficult for visitors, who due to social norms cannot control devices owned by hosts to assess whether a microphone is, indeed, 'off' upon visual inspection only. Thus, we posit that in addition to using LED based feedback, devices need to incorporate feedback mechanisms that tangibly communicates the device's state to all the occupants in the vicinity.

In general, our findings indicate that people prefer on-device tangible control mechanisms to disable the microphone which is inline with prior work [102]. 'Showing' users a physical hardware disconnect did not appear to provide any additional trust than other mechanisms, indicating that users assume physical buttons are more reliable. Therefore a key design challenge for incorporating tangible control mechanisms in future smart voice assistants is that these buttons need to have reliable disconnects, in a way that is verifiable by either users or other experts. More thought is needed on how to provide strong assurances to users that a hardware disconnect has taken place. The lack of trust with current designs is the confusion over what these switches actually do [59].

### 6.2 Hardware Switches and The Control Paradox

We see evidence of the 'control paradox' in our findings. The control paradox refers to the phenomenon where people are willing to take on more risk or judge risks as less severe when they feel they are in control [93] (the feeling of being in control can obscure the underlying risks associated). For example, although driving is objectively riskier than flying, people consider driving safer than flying because they feel more in control while driving a car. In the context of privacy, when people perceive more control over the release of private information they are more likely to share information with others [14, 76].

In line with the control paradox, our results show a negative correlation between perceived control and risk perception of the prototypes. One interpretation of this observation relates to the effect of uncertainty on risk perception. Uncertainty is considered as a key construct of risk [10, 92]; with the provided on-device tangible physical controls, our participants might have felt less uncertainty around the outcomes of their actions (e.g., muting the microphone with high assurance), hence they were more likely to perform the actions – disregarding the risk of negative outcomes. Prior studies has shown the lack of control points to lack of certainty in the context of behavioral trust [35]. In our context, providing tangible control mechanisms might have reduced the existing uncertainty/ambiguity about a prototype's state (whether it is actually off) which is considered to be a key property of tangible privacy [5], and hence, our participants might have considered the prototypes to be less risky.

On the one hand, it may be the case that improved controls provide enhanced privacy by decreasing uncertainty of when the device can and cannot record. At the minimum, therefore, devices should be designed so that the controls match the users' mental models of the controls – a tangible switch should reliably mute the microphone. On the other hand, with reliably designed controls, users might take on more risk and use these devices in unsafe ways (e.g., leaving them unmuted in sensitive environments and situations). Future studies will need to examine such possible behaviors and the resulting implications.

### 6.3 Tangible Control Mechanisms: Both Usable *and* Privacy Enhancing

We find an interesting relationship between usability and hardware based tangible control mechanisms. There is evidence that providing more attention to privacy and security generally results in decreased usability [42]. However, in our study, we found that switching from software based control mechanisms to hardware based mechanisms not only significantly *increased* usability but also enhanced participants' sense of privacy in the context of eavesdropping. In particular, we found that the use of physical control mechanisms significantly reduced participants' perceptions of the risk of being eavesdropped, showing the use of tangible control mechanisms does not necessarily represent a trade-off.

Another important implication of our study is its contribution towards the Privacy by Design (PbD) approach [20] for smart voice assistants. As IoT devices are increasingly becoming more ubiquitous, PbD has been advocated and recommended by several studies as a means of preserving user privacy [57, 74]. Prior research has proposed several frameworks to integrate PbD into the design space of IoT devices [3, 43, 79]. Our study advances this body of work by designing prototypes that are in line with the key recommendations of PbD. In particular, tangible control mechanisms provide *proactive* privacy measures that can be *embedded into the design space* of voice assistants [20, p. 2-3]. Moreover, assured feedback mechanisms provide *visibility and transparency* about the effect of the control mechanisms to the users [20, p. 4]. Overall, we hope our prototypical designs and the findings of our statistical analysis will nudge device manufacturers towards considering hardware based mechanisms (beyond software based mechanisms) for providing PbD in the context of smart voice assistants.

### 6.4 Limitations

Virtual prototypes [98] are an important step in the design process. As mentioned earlier, virtual prototyping has been shown to be a valid method for assessing the usability of potential designs (including tangible design prototypes) [21]. They also provides savings in time and cost, especially in the context of avoiding burdensome costs for hardware and software development earlier in the design process [21, 98]. Nevertheless, virtual prototypes do not recreate the entire experience and thus 'final' designs would still need to be studied through hardware prototypes. Thus our results should be considered an important step in the design of tangible privacy interfaces, but not as the final word. Our findings showing that physical control mechanisms are privacy enhancing should be interpreted in the context of participants' risk perceptions of the prototypes (i.e., applying physical control mechanisms makes the prototypes feel less risky to people in terms of eavesdropping). Privacy as a larger concept encompasses a range of phenomenological aspects [70] and our dependent variables do not capture all the aspects of privacy.

Our participants were crowd workers recruited on Amazon Mechanical Turk [17]. Although prior works suggest that data collected from Amazon MTurk is reflective of real-world behavior in different contexts [32] we acknowledge that the population is not representative of the U.S.; studies have found that MTurkers differs from the U.S. population in terms of their age, education, and privacy concerns [53, 81]. Although usability testing was performed online, and the quality of results from tests on MTurk might not be as good as lab based testing, prior work has validated online crowd-based approaches as a viable platform for usability testing [61]. Finally, our focus on the U.S. population also means that our results may not generalize to other countries, and further work is needed to understand how other cultures interpret tangible designs.

## 7 CONCLUSIONS

Although people can be assured of their privacy by covering or obscuring cameras, no clear solutions exist to assure people that embedded microphones are indeed 'off'. Prior work has suggested that such 'tangible privacy' would need physical controls as well as assured feedback that the microphone was indeed disconnected. We conducted a user study (N=261) to evaluate various design combinations of smart voice assistants, offering hardware vs. software controls and different approaches (e.g., showing the physical disconnection of the microphone) for assured feedback about the microphone's state.

Our first major finding is that a tangible, physical control to mute/unmute devices provides a statistically significant increase in people's perceptions of reliability, trust (and reduced risk), usability, and control of the device. Thus it seems clear that *future designs of smart voice assistants should use a tangible control, i.e., a hardware based switch, to mute the microphone.*

Our second major finding is that, interestingly, the tested feedback mechanisms, i.e., how the devices communicate their on/off state to the users, did not have any statistically significant influence on any of the constructs. Thus it appears that users are trusting of hardware based switches and even *showing* the physical disconnection of the microphone was not more trustworthy than software or LED indicators. Thus, a major design implication is that a large amount of responsibility lies on device manufacturers. *Device manufacturers need to ensure their designs do* *indeed disconnect the microphones at a hardware level* to be in line with people's expectations, since this functionality is assumed. Explanations from our survey participants indicate a preference for physical feedback (e.g., seeing the microphone get disconnected) and LED based feedback, so these mechanisms should not be removed; even though one may not be preferred over the other, feedback is expected. In general, the use of software applications (for control and feedback) is perceived as unreliable and untrustworthy by our participants.

The third major contribution of our study is providing the research community with quantitative evidence regarding the effectiveness of tangible privacy mechanisms in protecting *both* user privacy and enhancing usability in the context of smart voice assistants with embedded microphones. Thus, *tangible privacy mechanisms do* not *represent a usability-privacy trade-off.*

In conclusion, we recommend that smart voice assistants should provide users with tangible (physical) mechanisms for controlling microphones and note that device manufacturers are entrusted to disconnect the microphone at the physical level. However, further studies are needed to more suitably convey feedback about the connection state to users. Although some do find the approach useful, showing users the disconnected microphone did not appear to improve overall trust any more than LED or software based approaches, and thus there is room for exploring other novel approaches that apply more generally.

## REFERENCES

[1] Noura Abdi, Kopo M. Ramokapane, and Jose M. Such. 2019. More than Smart Speakers: Security and Privacy Perceptions of Smart Home Personal Assistants. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. USENIX Association, Santa Clara, CA, 451–466. https://www.usenix.org/conference/soups2019/presentation/abdi

[2] Noura Abdi, Xiao Zhan, Kopo M. Ramokapane, and Jose Such. 2021. Privacy Norms for Smart Home Personal Assistants. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, Article 558, 14 pages. https://doi.org/10.1145/3411764.3445122

[3] Hezam Akram Abdulghani, Niels Alexander Nijdam, Anastasija Collen, and Dimitri Konstantas. 2019. A study on security and privacy guidelines, countermeasures, threats: IoT data at rest perspective. *Symmetry* 11, 6 (2019), 774.

[4] Abbas Acar, Hossein Fereidooni, Tigist Abera, Amit Kumar Sikder, Markus Miettinen, Hidayet Aksu, Mauro Conti, Ahmad-Reza Sadeghi, and Selcuk Uluagac. 2020. Peek-a-boo: I see your smart home activities, even encrypted!. In *Proceedings of the 13th ACM Conference on Security and Privacy in Wireless and Mobile Networks*. 207–218.

[5] Imtiaz Ahmad, Rosta Farzan, Apu Kapadia, and Adam J. Lee. 2020. Tangible Privacy: Towards User-Centric Sensor Designs for Bystander Privacy. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2, Article 116 (Oct. 2020), 28 pages.

[6] Tawfiq Ammari, Jofish Kaye, Janice Y Tsai, and Frank Bentley. 2019. Music, Search, and IoT: How People (Really) Use Voice Assistants. *ACM Trans. Comput. Hum. Interact.* 26, 3 (2019), 17–1.

[7] Noah Apthorpe, Dillon Reisman, Srikanth Sundaresan, Arvind Narayanan, and Nick Feamster. 2017. Spying on the smart home: Privacy attacks and defenses on encrypted iot traffic. *arXiv preprint arXiv:1708.05044* abs/1708.05044 (2017).

[8] David Armstrong, Ann Gosling, John Weinman, and Theresa Marteau. 1997. The place of inter-rater reliability in qualitative research: An empirical study. *Sociology* 31, 3 (1997), 597–606.

[9] Katrin Arning and Martina Ziefle. 2007. Understanding age differences in PDA acceptance and performance. *Computers in Human Behavior* 23, 6 (2007), 2904–2927.

[10] Terje Aven and Ortwin Renn. 2009. On risk defined as an event where the outcome is uncertain. *Journal of risk research* 12, 1 (2009), 1–11.

[11] Nazmiye Balta-Ozkan, Oscar Amerighi, and Benjamin Boteler. 2014. A comparison of consumer perceptions towards smart homes in the UK, Germany and Italy: reflections for policy and future research. *Technology Analysis & Strategic Management* 26, 10 (2014), 1176–1195.

[12] G Beier. 1999. Locus of control when interacting with technology (Kontrollüberzeugungen im Umgang mit Technik). *Report Psychologie* 24, 9 (1999), 684–693.

[13] Karen Bonilla and Aqueasha Martin-Hammond. 2020. Older adults' perceptions of intelligent voice assistant privacy, transparency, and online privacy guidelines. In *Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020)*.

[14] Laura Brandimarte, Alessandro Acquisti, and George Loewenstein. 2013. Misplaced confidences: Privacy and the control paradox. *Social psychological and personality science* 4, 3 (2013), 340–347.

[15] John Brooke. 1996. Sus: a "quick and dirty'usability. *Usability evaluation in industry* 189 (1996).

[16] Joseph Bugeja, Andreas Jacobsson, and Paul Davidsson. 2016. On Privacy and Security Challenges in Smart Connected Homes. In *2016 European Intelligence and Security Informatics Conference (EISIC)*. 172–175. https://doi.org/10.1109/EISIC.2016.044

[17] Michael Buhrmester, Tracy Kwang, and Samuel D Gosling. 2016. Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data? 6 (2016).

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

[19] Nico Castelli, Corinna Ogonowski, Timo Jakobi, Martin Stein, Gunnar Stevens, and Volker Wulf. 2017. What Happened in My Home?: An End-User Development Approach for Smart Home Data Visualization. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. ACM, New York, NY, USA, 853–866.

[20] Ann Cavoukian. 2009. Privacy by design. (2009).

[21] S. Ceccacci and M. Mengoni. 2017. Designing Smart Home Interfaces: Traditional vs Virtual Prototyping. In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments* (Island of Rhodes, Greece) *(PETRA '17)*. Association for Computing Machinery, New York, NY, USA, 67–74. https://doi.org/10.1145/3056540.3056556

[22] Marie Chan, Daniel Estève, Christophe Escriba, and Eric Campo. 2008. A review of smart homes—Present state and future challenges. *Computer methods and programs in biomedicine* 91, 1 (2008), 55–81.

[23] Varun Chandrasekaran, Suman Banerjee, Bilge Mutlu, and Kassem Fawaz. 2021. 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)*. 535–552.

[24] Varun Chandrasekaran, Kassem Fawaz, Bilge Mutlu, and Suman Banerjee. 2018. Characterizing privacy perceptions of voice assistants: A technology probe study. *arXiv preprint arXiv:1812.00263* 135 (2018).

[25] Gary Charness and Uri Gneezy. 2012. Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization* 83, 1 (2012), 50–58.

[26] Hsuan-Ting Chen and Wenhong Chen. 2015. Couldn't or wouldn't? The influence of privacy concerns and self-efficacy in privacy management on privacy protection. *Cyberpsychology, Behavior, and Social Networking* 18, 1 (2015), 13–19.

[27] Yuxin Chen, Huiying Li, Steven Nagels, Zhijing Li, Pedro Lopes, Ben Y Zhao, and Haitao Zheng. 2019. Understanding the effectiveness of ultrasonic microphone jammer. *arXiv preprint arXiv:1904.08490* abs/1904.08490 (2019).

[28] Eugene Cho. 2019. Hey Google, can I ask you something in private?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–9.

[29] Eugene Cho, S Shyam Sundar, Saeed Abdullah, and Nasim Motalebi. 2020. Will deleting history make alexa more trustworthy? effects of privacy and content customization on user experience of smart speakers. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[30] Yaliang Chuang, Lin-Lin Chen, and Yoga Liu. 2018. Design Vocabulary for Human-IoT Systems Communication. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, Article 274, 11 pages.

[31] Hyunji Chung, Michaela Iorga, Jeffrey Voas, and Sangjin Lee. 2017. Alexa, can I trust you? *Computer* 50, 9 (2017), 100–104.

[32] Scott Clifford, Ryan M Jewell, and Philip D Waggoner. 2015. Are samples drawn from Mechanical Turk valid for research on political ideology? *Research & Politics* 2, 4 (2015), 2053168015622072.

[33] Sofia Elena Colesca. 2009. Understanding trust in e-government. *Engineering Economics* 63, 3 (2009).

[34] Ry Crist and Andrew Gebhart. 2018. Everything you need to know about the Amazon Echo. *CNET, September* 21 (2018).

[35] TK Das and Bing-Sheng Teng. 2004. The risk-based view of trust: A conceptual framework. *journal of Business and Psychology* 19, 1 (2004), 85–116.

[36] Nora A Draper and Joseph Turow. 2019. The corporate cultivation of digital resignation. *New Media & Society* 21, 8 (2019), 1824–1839.

[37] Kiran K Edara. 2014. Key word determinations from voice data. US Patent 8,798,995.

[38] Pardis Emami-Naeini, Janarth Dheenadhayalan, Yuvraj Agarwal, and Lorrie Faith Cranor. 2021. Which Privacy and Security Attributes Most Impact Consumers' Risk Perception and Willingness to Purchase IoT Devices?. In *2021 IEEE Symposium on Security and Privacy (SP)*. 1937–1954.

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

[40] Andy Field, Jeremy Miles, and Zoë Field. 2012. *Discovering statistics using R.* Sage publications.

[41] Morten Fjeld, Sissel Guttormsen Schar, Domenico Signorello, and Helmut Krueger. 2002. Alternative tools for tangible interaction: A usability evaluation. In *Proceedings. International Symposium on Mixed and Augmented Reality*. IEEE, 157–318.

[42] Ivan Fléchais. 2005. Designing secure and usable systems. *PhD diss., University College London* (2005).

[43] Noria Foukia, David Billard, and Eduardo Solana. 2016. PISCES: A framework for privacy by design in IoT. In *2016 14th Annual Conference on Privacy, Security and Trust (PST)*. IEEE, 706–713.

[44] Nathaniel Fruchter and Ilaria Liccardi. 2018. Consumer attitudes towards privacy and security in home assistants. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–6.

[45] Chuhan Gao, Varun Chandrasekaran, Kassem Fawaz, and Suman Banerjee. 2018. Traversing the quagmire that is privacy in your smart home. In *Proceedings of the 2018 Workshop on IoT Security and Privacy*. 22–28.

[46] Siddharth Gulati, Sonia Sousa, and David Lamas. 2018. Modelling trust in human-like technologies. In *Proceedings of the 9th Indian Conference on Human Computer Interaction*. 1–10.

[47] Pamela J Hinds. 1998. *User control and its many facets: A study of perceived control in human-computer interaction*. Hewlett Packard Laboratories California.

[48] Gary Horcher. 2018. Woman says her Amazon device recorded private conversation, sent it out to random contact. *KIRO News* (2018).

[49] Eva Hornecker and Jacob Buur. 2006. Getting a grip on tangible interaction: a framework on physical space and social interaction. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 437–446.

[50] Roberto Hoyle, Luke Stark, Qatrunnada Ismail, David Crandall, Apu Kapadia, and Denise Anthony. 2020. Privacy Norms and Preferences for Photos Posted Online. *ACM Transactions on Computer-Human Interaction (ACM TOCHI)* 27, 4, Article 30 (Aug. 2020), 27 pages. https://doi.org/10.1145/3380960

[51] Yue Huang, Borke Obada-Obieh, and Konstantin Beznosov. 2020. 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*. 1–13.

[52] Jiun-Yin Jian, Ann M Bisantz, and Colin G Drury. 2000. Foundations for an empirically determined scale of trust in automated systems. *International journal of cognitive ergonomics* 4, 1 (2000), 53–71.

[53] Ruogu Kang, Stephanie Brown, Laura Dabbish, and Sara Kiesler. 2014. Privacy attitudes of mechanical turk workers and the us public. In *10th Symposium On Usable Privacy and Security ({SOUPS} 2014)*. 37–49.

[54] Bjørn Karmann. 2018. Project Alias. http://bjoernkarmann.dk/project_alias. Accessed: 2021-06-30.

[55] RM Khan and MA Khan. 2007. Academic sojourners, culture shock and intercultural adaptation: A trend analysis. *Studies About Languages* 10 (2007), 38–46.

[56] Harri Kiljander et al. 2004. *Evolution and usability of mobile phone interaction styles*. Helsinki University of Technology.

[57] Demetrius Klitou. 2014. Privacy-invading technologies and privacy by design. *Information Technology and Law Series* 25 (2014), 27–45.

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

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

[60] Hannah Limerick, David Coyle, and James W Moore. 2014. The experience of agency in human-computer interactions: a review. *Frontiers in human neuroscience* 8 (2014), 643.

[61] Di Liu, Randolph G Bias, Matthew Lease, and Rebecca Kuipers. 2012. Crowdsourcing for usability testing. *Proceedings of the American Society for Information Science and Technology* 49, 1 (2012), 1–10.

[62] Yuchen Liu, Ziyu Xiang, Eun Ji Seong, Apu Kapadia, and Donald S. Williamson. 2021. Defending Against Microphone-Based Attacks with Personalized Noise. *Proceedings on Privacy Enhancing Technologies (PoPETs)* 2021, 2 (2021), 130–150. https://doi.org/10.2478/popets-2021-0021

[63] Maria Madsen and Shirley Gregor. 2000. Measuring human-computer trust. In *11th australasian conference on information systems*, Vol. 53. Citeseer, 6–8.

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

[65] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for CSCW and HCI practice. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–23.

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

[67] Adam W Meade and S Bartholomew Craig. 2012. Identifying careless responses in survey data. *Psychological methods* 17, 3 (2012), 437.

[68] Abraham H Mhaidli, Manikandan Kandadai Venkatesh, Yixin Zou, Florian Schaub, and M Kandadai. 2020. Listen Only When Spoken To: Interpersonal Communication Cues as Smart Speaker Privacy Controls. *Proc. Priv. Enhancing Technol.* 2020, 2 (2020), 251–270.

[69] Don Norman. 2014. *Turn signals are the facial expressions of automobiles.* Diversion Books.

[70] Kieron O'Hara. 2016. The seven veils of privacy. *IEEE Internet Computing* 20, 2 (2016), 86–91.

[71] Ilesanmi Olade, Christopher Champion, Haining Liang, and Charles Fleming. 2019. The $Smart^2$ Speaker Blocker: An Open-Source Privacy Filter for Connected Home Speakers. *arXiv preprint arXiv:1901.04879* abs/1901.04879 (2019).

[72] Jason W Osborne, Anna B Costello, and J Thomas Kellow. 2014. *Best practices in exploratory factor analysis.* CreateSpace Independent Publishing Platform Louisville, KY.

[73] Antti Oulasvirta, Aurora Pihlajamaa, Jukka Perkiö, Debarshi Ray, Taneli Vähäkangas, Tero Hasu, Niklas Vainio, and Petri Myllymäki. 2012. Long-term effects of ubiquitous surveillance in the home. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing.* 41–50.

[74] Yvonne O'Connor, Wendy Rowan, Laura Lynch, and Ciara Heavin. 2017. Privacy by design: informed consent and internet of things for smart health. *Procedia computer science* 113 (2017), 653–658.

[75] Yin Minn Pa Pa, Shogo Suzuki, Katsunari Yoshioka, Tsutomu Matsumoto, Takahiro Kasama, and Christian Rossow. 2015. IoTPOT: Analysing the rise of IoT compromises. In *9th {USENIX} Workshop on Offensive Technologies ({WOOT} 15).*

[76] Sameer Patil, Roman Schlegel, Apu Kapadia, and Adam J. Lee. 2014. Reflection or Action?: How Feedback and Control Affect Location Sharing Decisions. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '14).* 101–110. https://doi.org/10.1145/2556288.2557121

[77] Joann Peck and Terry L Childers. 2003. To have and to hold: The influence of haptic information on product judgments. *Journal of Marketing* 67, 2 (2003), 35–48.

[78] Joann Peck and Jennifer Wiggins Johnson. 2011. Autotelic need for touch, haptics, and persuasion: The role of involvement. *Psychology & Marketing* 28, 3 (2011), 222–239.

[79] Charith Perera, Ciaran McCormick, Arosha K Bandara, Blaine A Price, and Bashar Nuseibeh. 2016. Privacy-by-design framework for assessing internet of things applications and platforms. In *Proceedings of the 6th International Conference on the Internet of Things.* 83–92.

[80] Qualtrics. 2021. Qualtrics XM. https://www.qualtrics.com/. Accessed: 2021-06-10.

[81] Joel Ross, Lilly Irani, M Six Silberman, Andrew Zaldivar, and Bill Tomlinson. 2010. Who are the crowdworkers? Shifting demographics in Mechanical Turk. In *CHI'10 extended abstracts on Human factors in computing systems.* 2863–2872.

[82] Mary Beth Rosson and John M Carroll. 2009. Scenario-based design. In *Human-computer interaction.* CRC Press, 161–180.

[83] Matthias Rothensee. 2008. User acceptance of the intelligent fridge: empirical results from a simulation. In *The internet of things.* Springer, 123–139.

[84] Yosiyuki Sakamoto, Makio Ishiguro, and Genshiro Kitagawa. 1986. Akaike information criterion statistics. *Dordrecht, The Netherlands: D. Reidel* 81, 10.5555 (1986), 26853.

[85] Mike Scaife and Yvonne Rogers. 1996. External cognition: how do graphical representations work? *International journal of human-computer studies* 45, 2 (1996), 185–213.

[86] Kim Bartel Sheehan. 1999. An investigation of gender differences in on-line privacy concerns and resultant behaviors. *Journal of interactive marketing* 13, 4 (1999), 24–38.

[87] Kim Bartel Sheehan. 2002. Toward a typology of Internet users and online privacy concerns. *The information society* 18, 1 (2002), 21–32.

[88] Dong-Hee Shin. 2010. The effects of trust, security and privacy in social networking: A security-based approach to understand the pattern of adoption. *Interacting with computers* 22, 5 (2010), 428–438.

[89] Dong-Hee Shin. 2012. Cross-analysis of usability and aesthetic in smart devices: what influences users' preferences? *Cross Cultural Management: An International Journal* 19 (2012), 563–587.

[90] Ben Shneiderman and Catherine Plaisant. 2010. *Designing the user interface: Strategies for effective human-computer interaction.* Pearson Education India.

[91] Vijay Sivaraman, Hassan Habibi Gharakheili, Arun Vishwanath, Roksana Boreli, and Olivier Mehani. 2015. Network-level security and privacy control for smart-home IoT devices. In *2015 IEEE 11th International conference on wireless and mobile computing, networking and communications (WiMob).* IEEE, 163–167.

[92] Lennart Sjöberg, Bjørg-Elin Moen, and Torbjørn Rundmo. 2004. Explaining risk perception. *An evaluation of the psychometric paradigm in risk perception research* 10, 2 (2004), 665–612.

[93] Paul Slovic. 1987. Perception of risk. *Science* 236, 4799 (1987), 280–285.

[94] Madiha Tabassum, Tomasz Kosiński, Alisa Frik, Nathan Malkin, Primal Wijesekera, Serge Egelman, and Heather Richter Lipford. 2019. Investigating Users' Preferences and Expectations for Always-Listening Voice Assistants. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4 (2019), 1–23.

[95] Madiha Tabassum, Tomasz Kosinski, and Heather Richter Lipford. 2019. " I don't own the data": End User Perceptions of Smart Home Device Data Practices and Risks. In *Fifteenth Symposium on Usable Privacy and Security ({SOUPS} 2019)*. 435–450.

[96] Brygg Ullmer and Hiroshi Ishii. 2000. Emerging frameworks for tangible user interfaces. *IBM systems journal* 39, 3.4 (2000), 915–931.

[97] Elise Van Den Hoven, Evelien Van De Garde-Perik, Serge Offermans, Koen Van Boerdonk, and Kars-Michiel H Lenssen. 2013. Moving Tangible Interaction Systems to the Next Level. *IEEE Computer* 46, 8 (2013), 70–76.

[98] G Gary Wang. 2002. Definition and review of virtual prototyping. *J. Comput. Inf. Sci. Eng.* 2, 3 (2002), 232–236.

[99] Consumer Watchdog. 2017. Google, Amazon Patent Filings Reveal Digital Home Assistant Privacy Problems.

[100] Stephan AG Wensveen, Johan Partomo Djajadiningrat, and CJ Overbeeke. 2004. Interaction frogger: a design framework to couple action and function through feedback and feedforward. In *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques*. 177–184.

[101] Heng Xu, Tamara Dinev, H Jeff Smith, and Paul Hart. 2008. Examining the formation of individual's privacy concerns: Toward an integrative view. 6 (2008).

[102] Yaxing Yao, Justin Reed Basdeo, Smirity Kaushik, and Yang Wang. 2019. 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*. 1–12.

[103] An Gie Yong, Sean Pearce, et al. 2013. A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in quantitative methods for psychology* 9, 2 (2013), 79–94.

[104] Eva-Maria Zeissig, Chantal Lidynia, Luisa Vervier, Andera Gadeib, and Martina Ziefle. 2017. Online privacy perceptions of older adults. In *International Conference on Human Aspects of IT for the Aged Population*. Springer, 181–200.

[105] Eric Zeng, Shrirang Mare, and Franziska Roesner. 2017. End user security and privacy concerns with smart homes. In *thirteenth symposium on usable privacy and security ({SOUPS} 2017)*. 65–80.

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

[107] Oren Zuckerman. 2015. Objects for change: A case study of a tangible user interface for behavior change. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction*. 649–654.

## A APPENDIX

### A.1 Usability Questionnaire

- **U1:** I would frequently use this approach to mute the microphone.
- **U2:** I found this approach to mute the microphone unnecessarily complex.
- **U3:** I think this approach to mute the microphone is easy to use.
- **U4:** I would need the support of a technical person to be able to use this approach to mute the microphone.
- **U5:** I found the various functions/parts to mute the microphone work well together.
- **U6:** I found there was too much inconsistency in various function/parts in this approach of muting the microphone.
- **U7:** I believe that most people would learn to use this approach to mute the microphone very quickly.
- **U8:** I found this approach to mute the microphone very cumbersome to use.
- **U9:** I would feel very confident when using this approach to mute the microphone.
- **U10:** I would need to learn a lot of things before I could get going with this approach to mute the microphone.

### A.2 Perceived Reliability Questionnaire

- **PR1:** By just looking at the device, I can tell whether the microphone is muted or unmuted.
- **PR2:** This is a reliable approach of controlling (muting/unmuting) the microphone.

- **PR3:** Using this approach to mute the microphone, I can rely on the device to not record me.

### A.3 Trust Questionnaire

- **T1:** This approach of muting the microphone is deceptive and misleading, it can lead me to believe that the device is not recording while it's actually recording.
- **T2:** Even if I mute the microphone using this approach, the device could behave in an underhanded manner by recording unwanted audio.
- **T3:** Even if I mute the microphone using this approach, I would be suspicious of the device's actions (whether it is recording or not).
- **T4:** I am skeptical of the visual information provided by this approach of muting the microphone.
- **T5:** I am confident that this approach mutes the microphone and prevents unwanted audio recording.
- **T6:** This approach of muting the microphone provides security against unwanted audio recording.
- **T7:** This approach of muting the microphone is dependable in preventing unwanted audio recording.
- **T8:** I can trust this approach of muting the microphone for preventing unwanted audio recording.

### A.4 Perceived Control Questionnaire

- **C1:** With this approach of muting the microphone, I feel I am in control of any recordings by this device.
- **C2:** With this approach of muting the microphone, I am able to stop audio recording on my own without relying on the device to do so.

### A.5 Risk Perception Questionnaire

- **R1:** Even if I mute the microphone using this approach, I would feel unsafe having a private conversation around this device.
- **R2:** Even if I mute the microphone using this approach, I feel there could be negative consequences (e.g. unintended recording) while having a conversation around this device.
- **R3:** When I mute the microphone using this approach, I feel I must be cautious when having a private conversation around this device.
- **R4:** When I mute the microphone using this approach, I feel it would be risky to have a private conversation around this device.

### A.6 Technical Competence Questionnaire

- Usually, I successfully cope with technical problems.
- Even if I face problems while coping with technical problems, I continue working on them.
- I really enjoy cracking technical problems.
- Up to now I managed to solve most of the technical problems, therefore I am not afraid of them in future.
- I better keep my hands off technical devices because I feel uncomfortable and helpless about them.
- Technical devices are often not transparent (i.e., easy to understand) and difficult to handle.
- When I solve a technical problem successfully, it mostly happens by chance.
- Most technical problems are too complicated to deal with them.

### A.7 Privacy Awareness Questionnaire

- I follow the news and developments about privacy issues and privacy violations.
- I cannot comprehend the relevance of the issue of privacy because I do not care about it.
- I pay closer attention to privacy issues and privacy violations since they have become so prominent in the media.

### A.8 Privacy Self-efficacy Questionnaire

- I know most privacy settings of the devices I use.
- Because I have had no problems with privacy settings so far, I am confident for future privacy tasks.
- I do not read privacy policies because I do not understand them.
- I always change my privacy settings when I start using a new device.
- I always change my privacy settings when I start using a new app.
- I feel helpless with privacy settings and measures, so I do not change anything.

### A.9 Privacy Preference Question

- Are you a private person who keeps to yourself or an open person who enjoys sharing with others? 1) Very private . . . 7) Very open.

### A.10 Open Ended Questions

- Did you find this approach of muting the microphone easy to use (or not)? Briefly explain why.
- Do you think this approach is a reliable (or not reliable) way of preventing unwanted audio recordings? Briefly explain why.
- Briefly explain some of the features of this approach of muting the microphone that you liked (or disliked).

### A.11 Underlying Factors: Privacy Confidence

- **PC1:** I know most privacy settings of the devices I use.
- **PC2:** Because I have had no problems with privacy settings so far,I am confident for future privacy tasks.
- **PC3:** Are you a private person who keeps to yourself or an open person who enjoys sharing with others?

### A.12 Underlying Factors: Privacy Vigilance

- **PV1:** I always change my privacy settings when I start using a new device.
- **PV2:** I always change my privacy settings when I start using a new app.

### A.13 Underlying Factors: Privacy Awareness

- **PA1:** I follow the news and developments about privacy issues and privacy violations.
- **PA2:** I pay closer attention to privacy issues and privacy violations since they have become so prominent in the media.

### A.14 Underlying Factors: Privacy Comprehension

- **PComp1:** I do not read privacy policies because I do not understand them.
- **PComp2:** I feel helpless with privacy settings and measures, so I do not change anything.

- **PComp3:** I cannot comprehend the relevance of the issue of privacy because I do not care about it.

Table 5. Factor loading for scale items of Underlying Privacy Factors.

| Construct | Item | Factor Loading |
|---|---|---|
| Privacy Confidence | PC1 | 0.486 |
| | PC2 | 0.358 |
| | PC3 | 0.470 |
| Privacy Vigilance | PV1 | 0.952 |
| | PV2 | 0.866 |
| Privacy Awareness | PA1 | 0.863 |
| | PA2 | 0.832 |
| Privacy Comprehension | PComp1 | 0.353 |
| | PComp2 | 0.747 |
| | PComp3 | 0.681 |

Table 6. Factor loading for scale items of Dependent Variables.

| Construct | Item | Factor Loading |
|---|---|---|
| Perceived Usability | U1 | 0.591 |
| | U2 | -0.890 |
| | U3 | 0.861 |
| | U4 | -0.345 |
| | U5 | 0.744 |
| | U6 | -0.744 |
| | U7 | 0.687 |
| | U8 | -0.832 |
| | U9 | 0.471 |
| | U10 | -0.447 |
| Perceived Reliability | PR1 | 0.572 |
| | PR2 | 0.864 |
| | PR3 | 0.826 |
| Perceived Trust | T1 | -0.755 |
| | T2 | -0.864 |
| | T3 | -0.867 |
| | T4 | -0.842 |
| | T5 | 0.883 |
| | T6 | 0.825 |
| | T7 | 0.780 |
| | T8 | 0.918 |
| Perceived Control | C1 | 1.00 |
| | C2 | 1.00 |
| Risk Perception | R1 | 0.95 |
| | R2 | 0.87 |
| | R3 | 0.88 |
| | R4 | 0.87 |

Table 7. Regression Results (main effect) for perceived reliability, and risk perception

| | *Dependent variable:* | |
|---|---|---|
| | Perceived Reliability | Risk Perception |
| | (1) | (2) |
| technical competence | | $-0.218^{**}$ |
| | | (-0.434,-0.002) |
| | | $p = 0.049$ |
| privacy comprehension | $-0.214^{**}$ | $0.433^{**}$ |
| | (-0.424,-0.005) | (0.066,0.800) |
| | $p = 0.047$ | $p = 0.022$ |
| privacy confidence | $0.402^{***}$ | $-0.719^{***}$ |
| | (0.175,0.628) | (-1.038,-0.400) |
| | $p = 0.001$ | $p = 0.00002$ |
| privacy vigilance | $-0.149^{***}$ | $0.309^{***}$ |
| | (-0.241,-0.057) | (0.165,0.454) |
| | $p = 0.002$ | $p = 0.00004$ |
| privacy awareness | | $0.193^{**}$ |
| | | (0.014,0.373) |
| | | $p = 0.037$ |
| physical control | $0.424^{***}$ | $-0.489^{***}$ |
| | (0.264,0.584) | (-0.715,-0.264) |
| | $p = 0.00000$ | $p = 0.00003$ |
| physical feedback | 0.068 | $-0.125$ |
| | (-0.131,0.268) | (-0.406,0.155) |
| | $p = 0.504$ | $p = 0.382$ |
| LED feedback | 0.112 | $-0.013$ |
| | (-0.081,0.304) | (-0.283,0.257) |
| | $p = 0.259$ | $p = 0.926$ |
| Constant | $2.564^{***}$ | $2.702^{***}$ |
| | (2.014,3.115) | (1.335,4.069) |
| | $p = 0.000$ | $p = 0.0002$ |
| Observations | 261 | 261 |
| $R^2$ | 0.184 | 0.263 |
| Adjusted $R^2$ | 0.164 | 0.239 |
| Residual Std. Error | 0.652 (df = 254) | 0.913 (df = 252) |
| F Statistic | $9.516^{***}$ (df = 6; 254) | $11.231^{***}$ (df = 8; 252) |
| *Note:* | | $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$ |