Abraham Mhaidli*, Manikandan Kandadai Venkatesh, Yixin Zou, and Florian Schaub

Listen Only When Spoken To: Interpersonal Communication Cues as Smart Speaker Privacy Controls

Abstract: Internet of Things and smart home technologies pose challenges for providing effective privacy controls to users, as smart devices lack both traditional screens and input interfaces. We investigate the potential for leveraging interpersonal communication cues as privacy controls in the IoT context, in particular for smart speakers. We propose privacy controls based on two kinds of interpersonal communication cues – gaze direction and voice volume level – that only selectively activate a smart speaker's microphone or voice recognition when the device is being addressed, in order to avoid constant listening and speech recognition by the smart speaker microphones and reduce false device activation. We implement these privacy controls in a smart speaker prototype and assess their feasibility, usability and user perception in two lab studies. We find that privacy controls based on interpersonal communication cues are practical, do not impair the smart speaker's functionality, and can be easily used by users to selectively mute the microphone. Based on our findings, we discuss insights regarding the use of interpersonal cues as privacy controls for smart speakers and other IoT devices.

Keywords: Smart speaker, voice assistant, Internet of Things, privacy, privacy engineering.

DOI 10.2478/popets-2020-0026
Received 2019-08-31; revised 2019-12-15; accepted 2019-12-16.

***Corresponding Author: Abraham Mhaidli:** University of Michigan School of Information, E-mail: mhaidli@umich.edu
Manikandan Kandadai Venkatesh: University of Michigan School of Information, E-mail: kvmani@umich.edu
Yixin Zou: University of Michigan School of Information, E-mail: yixinz@umich.edu
Florian Schaub: University of Michigan School of Information, E-mail: fschaub@umich.edu

1 Introduction

Internet of Things (IoT) and smart home devices have gained considerable traction in the consumer market [4]. Technologies such as smart door locks, smart thermostats, and smart bulbs offer convenience and utility to users [4]. IoT devices often incorporate numerous sensors from microphones to cameras. Though these sensors are essential for the functionality of these devices, they may cause privacy concerns over what data such devices and their sensors collect, how the data is processed, and for what purposes the data is used [35, 41, 46, 47, 70]. Additionally, these sensors are often in the background or hidden from sight, continuously collecting information. Users may forget about the presence of IoT devices as well as their privacy settings, which can lead to unexpected privacy invasions [35, 36, 74].

An increasingly common category of smart home devices are smart speakers with voice assistants, such as Amazon Echo or Google Home. These speakers allow users to search for information online, order products, and control other IoT devices in the home [3]. Additionally, smart speakers can help users with tasks such as cooking [44], and keeping track of to-do lists [65].

To detect and process voice commands, a smart speaker's microphones actively listen for an activation keyword (e.g., "Alexa"). Once an activation keyword has been recognized, subsequent commands are streamed to the device manufacturer's backend infrastructure for interpretation and processing [7]. This typically means the device's microphone is perpetually active, always listening for an activation keyword, even when the user has no intention to interact with the device. This has potential privacy implications, such as accidental device activation, in which case sensitive audio could be transmitted to the manufacturer. Users may be off-put by the perceived loss of privacy [3, 41], and may thus refrain from purchasing a smart speaker and enjoying the utility it offers [37, 40, 43]. Moreover, recent smart speakers come equipped with additional sensors, such as video cameras and motion sensors (e.g., Amazon Echo Show, Facebook Portal), which may raise further privacy con-

cerns. Although smart speakers often have a mute button to manually deactivate the microphone, users find it cumbersome to constantly mute and unmute the device, and thus rarely make use of such privacy controls [37].

We study whether this tension between the need for continuous sensing and the burden of cumbersome privacy controls could be overcome by seamlessly integrating privacy controls into the user experience of the device. More specifically, we propose embedding privacy functionality into users' interaction with smart speakers by leveraging interpersonal communication cues to determine when the device should be active.

When people engage in conversations, they often naturally and instinctively indicate who they are addressing through non-verbal cues [19]. For instance, people tend to look at the person they are talking to in an in-person conversation [12, 19]. Similarly, people tend to whisper or speak quietly when they only want a person close-by to hear them [48]. Prior work has found that some users personify their smart speakers [56]; they may therefore be amenable to using similar interpersonal communication cues when interacting with the device. This suggests an opportunity to build more intuitive and usable privacy controls for smart speakers that leverage interpersonal cues. Specifically, we explore how interpersonal communication cues could be used to let a smart speaker recognize when it is being addressed, and only then activate its microphone. Rather than requiring users to manually adjust privacy settings (e.g., by pressing the mute button), privacy is controlled by dynamically limiting sensor activity based on contextual information [59, 61]. With this approach, the time the speaker's microphone is active could be reduced to only those situations in which the user actually intends to interact with the smart speaker, thus mitigating potential privacy concerns with constantly listening microphones and accidental device activation.

We explore this concept with two kinds of interpersonal communication cues: the user's voice-volume level and gaze direction. We created a functional smart speaker prototype, which allowed us to implement privacy controls for smart speakers leveraging gaze detection and voice volume. We evaluated the feasibility and usability of our privacy control approaches in two lab studies. Here, feasibility refers to whether the privacy controls work as intended without impairing the users' ability to activate the device when needed. – in other words, is it possible to build these controls in ways that do not hinder device activation. Usability refers to the ease, effectiveness, and comfort of using the proposed privacy controls. We find that both gaze and voice volume are feasible cues to control when a smart speaker's microphone should be on. This suggests that privacy controls based on interpersonal communication cues can be integrated into a smart speaker's user experience without impacting the smart speaker's utility. Furthermore, both privacy controls appear to be usable, i.e., our participants could easily and intuitively use these context-adaptive privacy controls. Participants further perceived the controls, in particular the voice volume-based privacy control, as useful and usable.

Based on our findings, we discuss design implications for privacy controls of smart speakers and IoT devices, as well as further opportunities and challenges for leveraging interpersonal communication cues to design context-adaptive and dynamic privacy controls.

2 Related Work

Smart speakers and other IoT technologies have raised privacy concerns and spurred research on privacy solutions. We first discuss related work on privacy concerns and perceptions regarding IoT and smart home devices, then review prior efforts addressing privacy issues.

2.1 IoT Privacy Concerns

IoT and smart home devices such as smart meters, thermostats and wearable fitness trackers are becoming increasingly common, yet the physical environment in which they are usually placed [14] and the sensitivity of types of data they collect [47] cause substantial privacy concerns. These concerns can inhibit their adoption by individuals [43]. For instance, household robots have been found to have concerning security vulnerabilities, e.g., attackers could easily leverage a built-in video camera to spy on users' activities in the bedroom [18].

Research has found that users generally desire strong privacy protection for IoT devices and request increased transparency about data practices, especially when data is collected in intimate spaces, when the types of data are sensitive, and the party with whom the data are shared is not perceived as benevolent [46, 47]. Butler et al. [10] found that end-users were concerned about the privacy and physical harm of teleoperated robots at home, yet they struggled to accurately anticipate privacy threats introduced by these robots. Similarly, Zeng et al.'s study [74] revealed a limited technical understanding of smart home technologies among users,

leading to incomplete mental models of potential threats and risky behaviors. This finding is echoed by Abdi et al. [1] and Malkin et al. [41], who found that users have limited knowledge of how smart speakers collect, store, and process data.

Tensions can further arise between multiple users with different levels of tech-savviness or control regarding the same device. For end-users of smart homes, power imbalance occurs when one household member is in charge of purchasing and setting up the device, hence having the opportunity to spy on other household members at ease [74]. Parents and teenagers may have conflicting expectations of what data should or should not be collected by home surveillance devices (e.g., Internet-connected locks and camera systems), which negatively impacts the trust relationships between them [67].

For smart speakers specifically, their microphones, which continuously listen for activation keywords, cause privacy concerns [21]. Prior research has shown that microphones are perceived as one of the most invasive and privacy-violating sensors [9, 53]. Fruchter and Liccardi's analysis of online reviews [27] identified key issues among users' privacy and security concerns related to smart speakers, such as the substantial amount and scope of collected data, and the perceived "creepiness" of these devices overhearing sensitive conversations. Similarly, Malkin et al.'s survey [41] revealed common concerns among smart speaker users about their recordings being used for advertising or being accessed by third-parties, even for benign purposes. Ammari et al.'s study also surfaces concerns about third-party use of voice recordings, as well as concerns about and experiences with random false activation of smart speakers [3]. Lau et al. [37] further differentiated privacy perceptions and concerns between smart speaker users and non-users through their diary and interview study; they found that non-users generally did not trust smart speaker companies or did not perceive the utility of smart speakers, whereas users hold fewer privacy concerns but still report tensions in multi-person households and with guests. While there is evidence of "privacy calculus," i.e., users reporting willingness to trade off their privacy for the convenience and benefits provided by these devices [27, 37, 75], this is often paired with resignation over either having to accept privacy incursions or forgoing the technology's benefits completely, with little opportunity to selectively control a technology's features and privacy implications [20, 37]. Liao et al. confirm such differences in privacy attitudes between voice assistant users and non-users [40].

2.2 Smart Speaker Privacy Controls

Today's smart speakers have several types of privacy controls: activation word, mute button, audio logs, speaker recognition, among others. Smart speakers listen for an *activation keyword* (e.g., "Alexa"). Audio is processed locally until the activation keyword is recognized, at which point subsequent audio is recorded and streamed to the device manufacturer's backend for extraction of voice commands [5]. Some smart speakers have a physical *mute button* which can be used to deactivate the microphone when desired. Through companion apps and websites, smart speaker users can access and delete past voice recordings [16]. Moreover, Amazon recently launched new voice commands as controls for the deletion of recorded voice commands, such as "delete everything I said today" and "delete what I just said" [33].

However, current privacy controls for smart speakers are problematic. The activation keyword approach is susceptible to false activation [3, 24], i.e., the smart speaker is activated when the user had no desire to activate the device. Few smart speaker users review or delete their recordings stored by the device's manufacturer, and many do not even know such option exists [3, 41]. Similarly, the mute button is rarely used [3, 37], as it requires a shift in modality [60] compared to normal interaction with the smart speaker (going to the device and pressing a button versus speaking to the device [37]). Furthermore, in situations in which a user wants to use the smart speaker for one task (e.g., controlling music playback) while simultaneously having a private conversation with another person, the mute button fails because the device cannot be attentive to requests (e.g., skipping a song) without also actively listening to the user's conversation. Even if processing is done locally, people might still be uncomfortable with having the microphone on in such situations.

Recent work has investigated privacy solutions for smart speakers and IoT devices. Chandrasekaran et al. [11] proposed *bolt-on* privacy-preserving interventions, which required the user to deploy a remote control to mute the smart speaker by cutting its power or jamming its microphone. Karmann [32] built an app that activates the smart speaker using a word or sound defined by the user, and interrupts the microphone otherwise. Other solutions include continuous authentication [25], deploying sounds that can be recorded by microphones but remain inaudible to humans [57, 58], using ultrasound jamming to address stealthy recording [28], providing an acoustic tagging device that helps

users to track conversation recording in the back-end system [13], processing voice at the edge without sending the voice commands to the cloud [15], and designing privacy indicators that more effectively signal a device's activity and data collection [22, 46, 55].

Context-aware computing principles [63, 64] have also been integrated into privacy research, in order to create systems and devices that dynamically adapt to changes in context based on available information [50, 59, 61]. For example, Schaub et al. [62] developed a calendar display that dynamically changes its content depending on the people present. Researchers have also developed solutions to dynamically adjust mobile app permissions depending on contextual information (e.g., where the user is located, or whether the smartphone is locked or not) [52, 72]. Similar context-aware mechanisms have been explored in smart home settings, e.g., to selectively exclude private photos from a slideshow in presence of other people [34]. We make novel contributions by investigating the application of context awareness for smart speaker privacy controls, specifically examining the detection of when a user wants to speak to a smart speaker, and only then activating the smart speaker's microphone.

3 Leveraging Interpersonal Cues for Privacy

Interpersonal cues, both verbal and nonverbal, play a important role in helping people understand and navigate the dynamics of human communication [17]. Language is a typical example of verbal cues, whereas body language (e.g., gestures, facial expressions, eye contact) and haptics are examples of nonverbal cues [31]. People subconsciously leverage interpersonal cues to transmit meaning and communicate effectively in order to achieve personal and social goals [19]. Interpersonal cues are often culture-specific, meaning that how they are used and interpreted varies across cultures (e.g., shaking one's head can be a gesture of rejection or affirmation depending on the culture [45]), but in a specific sociocultural context there is generally a consensus over a cue's meaning [19]. Interpersonal cues are also used to indicate if someone is participating in a conversation: subtle gestures such as who we look at (or don't look at) whilst conversing indicates who we are addressing or who we want (or don't want) to hear us [12, 19].

Smart speakers are voice-driven, meaning users talk to, and are talked to by, the smart speaker's voice assistant. Users anthropomorphize smart speakers and their voice assistants [56] likely because of the conversational nature of interaction and referring to the device with a personified name (e.g., "Alexa"). This personification of smart speakers suggests that using interpersonal cues in interactions with smart speakers may be intuitive and easy to learn for smart speaker users.

Therefore, given that (1) users are familiar with using interpersonal communication cues that indicate who should be participating in a conversation, and (2) smart speaker interaction is conversational, we propose to use certain interpersonal cues to indicate to a smart speaker when it should be listening for activation keywords and when not. We explore two interpersonal cues specifically: *speaking up to address the device*, while speaking quietly to have a private conversation; and *looking at the device* to indicate that it is being addressed. These are among the most common cues used in interpersonal conversations [26] and seem amenable to be adopted as privacy controls for smart speakers. These approaches are described in more detail in Sections 3.1 and 3.2.

Using interpersonal cues for smart speakers' privacy controls means that the smart speaker's microphone or voice recognition can be switched off most of the time, substantially reducing the amount of time the device could potentially listen to the user's conversations, e.g., due to false activation. In contrast to a physical mute button, leveraging interpersonal cues means that privacy control functionality can be integrated into the user's interaction with the device rather than requiring a modality shift (e.g., pressing the mute button) and only require minor changes to how the user interacts with the device. Furthermore, with this approach a smart speaker can both be attentive to when the user wants to make a request, while respecting the privacy of the user's conversations happening in parallel.

3.1 Regulating Privacy with Voice Volume

The volume level of speech has been shown to play many roles in interpersonal communication [71]. Speaking loudly can communicate anger and aggression [54]. In conversation, matching the volume of conversational partners can increase social desirability [49]. The pitch and energy of speech are also reliable indicators of emotional conversation [73]. Of particular interest for privacy is how voice volume levels are used to indicate the sensitivity of a conversation: people generally speak in lower volumes when talking about confidential topics they do not want others to overhear [48]. Conversely,

people speak loudly when they want to broadcast information and be heard by others.

Voice is already a common method for controlling IoT devices, but voice volume has typically not been considered so far. In addition to general voice interaction, other features of voice, such as pitch and tone [23], or non-verbal manifestations (e.g., humming) [66] have been studied to control cursors [66], video games [30], and musical instruments [23], often aiming to enhance accessibility or usability.

As voice volume indicates a conversation's sensitivity and the extent of one's desire to be heard, we can leverage this to control privacy aspects of a system. A privacy control based on voice volume would dynamically activate or deactivate a smart speaker's speech recognition functionality based on how loudly or quietly the user is speaking. This could be accomplished with a loudness sensor (i.e., a decibel meter or low-grade microphone) rather than a microphone capable of recording speech. When the user wants to address the smart speaker, they would speak slightly louder than normal, akin to calling out for someone. Only when the voice volume level surpasses a defined threshold would the smart speaker's speech recognition be activated to enable listening for the activation keyword. If the user does not intend to address the smart speaker, they can talk in normal volume, or at a quieter level, depending on how the activation threshold is set. In that case, the activation threshold would not be reached and speech recognition would remain inactive.

A potential challenge in realizing this approach is the definition of a volume level activation threshold that allows private conversations at a reasonable volume, without requiring the user to yell at the smart speaker when wanting to use it. Furthermore, distinguishing between conversation and other ambient noise based on volume level may be challenging. Very loud environments may accidentally activate the smart speaker's speech recognition functionality; so might people having a loud argument. However, a voice-volume based privacy control approach would still constitute a substantial improvement over the status quo: current smart speakers are always listening for the activation keyword. A voice-based privacy control would only activate the smart speaker's speech recognition functionality. Once speech recognition is activated, the smart speaker's activation keyword would still need to be detected before audio is recorded and transferred from the device to the manufacturer's backend for voice command recognition. Moreover, the activation threshold for voice volume could be customized to users' needs and their environments.

3.2 Regulating Privacy with Gaze

Another important interpersonal communication cue is eye contact. Eye contact and gaze direction play many roles in interpersonal communication [19, 26], such as indicating status, establishing what relationship exists between participants, and establishing closeness. Eye contact is used to determine who is participating in a conversation, and to what degree. When someone speaks, for example, they tend to look at the person they are talking to [12, 19]. Vertegaal et al. [69] found that gaze is a reliable and accurate measure to determine who is meant to be speaking and listening in a conversation, and leveraged this to create a conversational agent that more accurately estimates when it is being addressed and when it should respond. Gaze direction and eye contact have also been used to control IoT devices, such as lamps [42], household appliances [68], and smart devices that display email notifications to users [2].

We make a novel contribution by integrating eye contact into privacy mechanisms for smart speakers, giving users the agency to control their privacy through gaze. In our approach, when a user wants to address the smart speaker, they would look at the device. Upon recognizing that the user is facing the device, our gaze-based privacy control activates the smart speaker's microphone, thus enabling the smart speaker to listen for the activation keyword. Once the user moves their attention away from the smart speaker, i.e., they are no longer looking at the device, the smart speaker's microphone is deactivated again.

Such a gaze-based privacy control would require a camera sensor, which carries its own privacy concerns—camera sensors are perceived as particularly sensitive [38, 53]. However, this approach would be particularly suitable for smart speakers that already include a camera, such as the Amazon Echo Show[1] and Facebook Portal.[2] This camera could be leveraged to recognize faces and gaze direction, such as through head orientation or eye tracking, and thus determine when a person is looking at the smart speaker, at which point the microphone would be activated. To limit the privacy

1 https://www.amazon.com/All-new-Echo-Show-2nd-Gen/dp/B077SXWSRP

2 https://portal.facebook.com/

risks of our proposed system, image processing should be performed locally on the device and in real-time without recording the camera feed. The only information required is whether a person is looking at the device or not. Moreover, the camera can be isolated from all network connections, to minimize the risk of camera images being transmitted to external entities, and requiring an explicit command for other camera functionality involving video streaming, such as video calls.

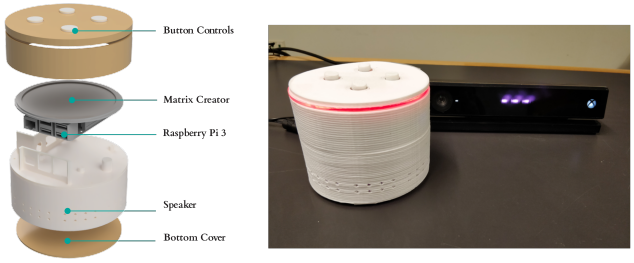

Fig. 1. Our smart speaker prototype: (a) schematic breakdown and (b) prototype with Kinect sensor used for voice-volume-based and gaze-based privacy controls.

3.3 Prototype

We realized the proposed voice-volume-based and gaze-direction-based privacy controls in a functional prototype, shown in Figure 1. Our prototype is based on a Rasberry Pi 3[3] with a Matrix Creator shield,[4] providing a microphone array. Using Sensory.ai[5] and Amazon Alexa Voice Services,[6] our smart speaker processes and responds to voice commands in the same way a commercial Amazon Echo device does. We 3D-printed a case for our prototype, so that our device resembles a smart speaker, including buttons to mute the microphone and to adjust the volume, modelled after Amazon Echo's buttons.

We used the Matrix Creator's ring of programmable LEDs to provide visual feedback about the device state. The LEDs are *red* when the speaker's microphone is deactivated. The LEDs are *green* when the speaker's microphone is active, signalling that the smart speaker is listening for the activation keyword. The LEDs are *blue* when the speaker's microphone is on *and* the voice assistant is engaged (i.e., the activation keyword has been recognized).

3.3.1 Voice volume detection

To implement the voice-volume-based privacy control, we used a second microphone instead of leveraging the Matrix Creator's microphones. This has multiple advantages: in contrast to the conventional activation keyword detection method, our approach decouples audio processing for the voice-volume-based privacy control from the smart speaker's general audio processing of voice commands. The voice-volume-based privacy control only needs to measure loudness (in decibels) in real time. No audio received by this second microphone needs to be ever recorded or analyzed in other ways. This decoupling of the two audio processing systems means that the privacy-related audio sensing and processing is self-contained and has low code complexity: it only compares the second microphone's loudness value against a pre-defined activation threshold. If the loudness is above the threshold, the smart speaker's primary microphone and speech recognition are activated, otherwise the primary microphone and speech recognition are deactivated.

In our device, loudness is measured as the energy of the audio signal. This is calculated as the root mean square value (RMS) of the audio signal represented within the range [–Min Energy, 0]. To determine the initial loudness threshold for our prototype, the authors engaged in informal pilot testing, speaking to the device at different volumes to activate it and identifying what volume level was easy enough to reach without straining users' voices but loud enough to clearly differentiate from private conversations. We determined -15 db (RMS) to be an appropriate threshold. This proved to be a reasonable threshold in our evaluation (see Sections 4 and 5). The threshold could be user-configurable to allow fine-tuning to different environments and user preferences.

The voice-volume-based privacy control integrates almost seamlessly into the smart speaker's user interaction, as shown in Figure 2. The only difference is that when a user wants to interact with the smart speaker they have to say the activation keyword (e.g., "Alexa") loudly enough, so that the decibel level is above the activation threshold. We use our prototype's LED ring to provide users with real-time visual feedback on loudness,

3 Raspberry Pi 3 Model B: https://www.raspberrypi.org/products/raspberry-pi-3-model-b/

4 Matrix Creator: https://www.matrix.one/products/creator

5 Sensory.ai Github repository: https://github.com/Sensory/alexa-rpi

6 Alexa Voice Services Github repository: https://github.com/alexa/alexa-avs-sample-app

i.e., the louder a user is speaking, the more LEDs light up, getting them closer or above the threshold, which is indicated on the ring by one blue LED.

When the activation threshold is reached, the smart speaker's primary microphone and speech recognition are activated. This activation is almost instantaneous, which means that the smart speaker's microphone is able to pick up enough of the activation keyword to still recognize it as such. In a performance evaluation study (see Section 4), we found the impact of this approach on smart speaker activation to be negligible. When the decibel level falls below the activation threshold, the primary microphone is deactivated. Other loud speech or ambient noise would also activate the primary microphone and speech recognition, but no command would be triggered unless the activation keyword is recognized. Note, that this corresponds to how smart speakers function ordinarily. Thus, our voice-volume-based privacy control adds an additional layer of privacy protection: users can now hold private conversations in vicinity of the smart speaker by speaking quietly; the smart speaker only listens for activation keywords when loudness is above the privacy threshold.

3.3.2 Gaze detection

To implement the gaze-based privacy control in our prototype, we used a Microsoft Kinect 2's depth-camera to recognize a user's head orientation, which functions as a proxy for gaze. We chose the Kinect as a readily avaialble off-the-shelf sensor that can track head orientation reliably. Eye tracking, though possible with the Kinect, proved to be too inaccurate to ensure reliable performance, but head tracking proved sufficient for our purposes. We use the camera sensor to determine whether a person is near the device and use head orientation (measured by pitch, yaw, and roll) to determine whether the user is looking at the device. If a person is looking at the device, the smartspeaker's microphone and speech recognition are activated; otherwise the microphone is deactivated. The head orientation is measured by the Kinect sensor in real time, no video images are recorded or stored.

Using the smart speaker with the gaze-based privacy control requires users to look at the device in order to indicate that they want to talk to the smart speaker. Triggering voice commands still requires the user to say the activation keyword ("Alexa"), as shown in Figure 2. While this constitutes a slight modification in how users interact with a smart speaker – requiring a line of sight between user and device – it effectively reduces the risk of false activation. The smart speaker would not be able to analyze audio unless the device is looked at, thus, preventing the device from misinterpreting snippets picked up from background conversations.

A caveat of the gaze-based privacy control is that it requires a camera sensor, which may be perceived as equally or even more intrusive as an always-listening microphone as shown in prior work [38, 53]. Therefore, this privacy control approach may be best suited for smart speakers that are already equipped with cameras to provide video-based functionality (e.g., Echo Show or Facebook Portal).

To evaluate the feasibility and usability of our privacy controls using interpersonal communication cues, we conducted two user studies. Our first user study assessed feasibility, i.e., whether and to what extent the addition of voice volume detection or gaze detection impacts users' ability to activate and interact with the smart speaker (see Section 4). Our second user study focused on usability, i.e., how comfortable participants were with using the proposed privacy controls, and whether they perceived the controls as having a positive (or negative) effect on their privacy (see Section 5). Both studies were approved by the University of Michigan's Institutional Review Board (IRB).

4 Study: Impact on Smart Speaker Activation

A concern in leveraging communication cues to determine the smart speaker's microphone state is that it may impact users' ability to activate and interact with the device. Our privacy controls may introduce a delay before the speaker's microphone is activated, or users might start speaking before they perform the cue (speaking up or looking at the device). We conducted a within-subjects lab experiment, in which we assessed the activation success rate of the voice-volume privacy control and the gaze-based privacy control. To establish a baseline, we further tested activation success rate of the same smart speaker prototype without dynamic privacy controls and of an Amazon Echo Dot.

Our findings show no significant difference in activation performance between the gaze-based or voice-volume-based privacy controls compared to our smart speaker prototype without those controls, which indicates that the effect of the proposed privacy controls on activation keyword detection is negligible, demonstrat-

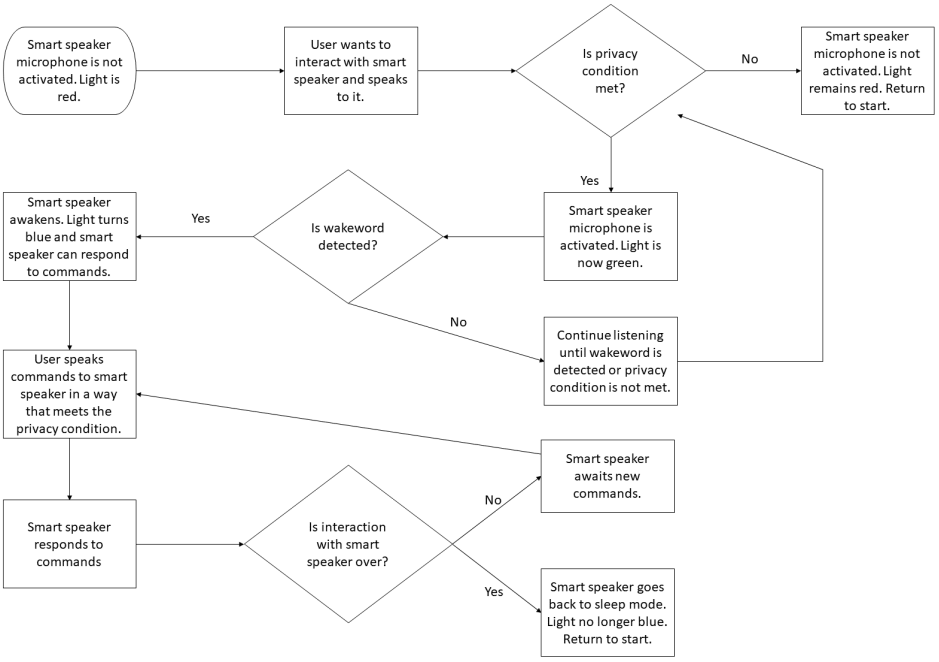

Fig. 2. Interaction diagram for device activation with the two proposed privacy controls. *Privacy condition* refers to either speaking above the volume threshold (for the voice-based control) or looking at the device (for the gaze-based control). Any commands spoken to the smart speaker must adhere to the respective privacy condition (i.e., commands must be spoken loudly in the Voice condition and while looking at the device for Gaze) else the primary microphone is deactivated and the smart speaker will not hear the commands.

ing the technical feasibility of the proposed controls. While the commercial smart speaker's activation performance was better, this was likely due to better microphones and audio processing logic unrelated to the proposed privacy controls.

4.1 Study Design and Setup

We evaluated activation performance in a within-subjects lab experiment with four conditions: our prototype with voice-volume-based privacy control (Voice), our prototype with gaze-based privacy control (Gaze), our prototype without the adaptive privacy controls (Control), and an Amazon Echo Dot (Echo Dot). Our smart speaker without privacy controls functioned as an internal control condition, and the Amazon Echo Dot functioned as an external control condition. All four conditions performed activation keyword detection locally using the same keyword ("Alexa"). However, the hardware and software used to detect the keyword differed between the prototype conditions and the Echo Dot. Our prototype used Sensory.ai and the Matrix Creator microphones; the Echo Dot used its proprietary hardware and software.

We tested the four conditions in two phases in which we carefully controlled multiple variables (e.g., distance and orientation to the device) in order to reliably assess and compare activation performance (see Figure 3). In the first phase, participants were introduced to the Voice, Control, and Echo Dot conditions, and asked to familiarize themselves with the mechanism. Once ready, participants stood at one of three distances (1.5m, 3m, and 4.5m) and were instructed to activate the smart speaker and have it flip a coin by saying "Alexa, flip a coin." We recorded the number of activation attempts to assess activation performance, measured as the number of times a participant had to say "Alexa, flip a coin" (or a variant thereof) before the device responded successfully (by flipping a coin and reporting the outcome).

The order in which a participant tested the conditions and the order of distances were randomized for each participant. Participants completed the task for the Voice, Control, and Echo Dot conditions at the same distance before moving to the next distance. We chose this approach to provide participants with a consistent context for modulating their voice for a given distance rather than having them frequently move among distances. After completing all distances, participants repeated the cycle two more times (thus doing each dis-

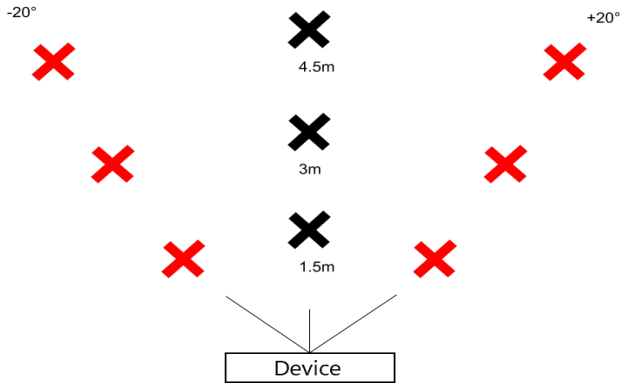

Fig. 3. Experimental setup for the performance evaluation. Participants stood at each X (1.5m, 3m, or 4.5m from the device) and attempted to activate the device. The red Xs are angled 20 degrees from the device, and were only used for gaze evaluation.

tance 3 times). This resulted in 9 data points (3 repetitions × 3 distances) per condition for each participant.

In the second phase, participants were introduced to the Gaze condition and asked to familiarize themselves with the mechanism. We measured participants' activation performance at different distances and angles away from the device with different body orientations, as these are aspects that might affect the reliability of gaze detection. There was a total of three distances (1.5m, 3m, and 4.5m), three angles per distance (0°, 20°, -20° away from the device), and three body orientations per angle, with feet and head facing forward, left (i.e., at a 90° angle from the device), or right (i.e., at a 90° angle from the device). Participants were instructed to stand at one of the positions with one of the body orientations, activate the smart speaker and have it flip a coin ("Alexa, flip a coin"). Participants were not allowed to move their feet, but could move their head, torso, and any other part of their body. After completing all distances, participants repeated the cycle two more times (thus doing each distance 3 times). This yielded 81 data points (3 repetitions × 3 distances × 3 angles × 3 orientations) for each participant in the Gaze condition.

We made the decision to randomize the voice-based conditions and test the Gaze condition afterwards to strike a trade-off between a fully randomized experiment and the complexity for participants in adjusting to different tasks, in favor of ensuring consistent context for the three voice conditions.

Each session lasted about 1 hour, and participants were compensated $15 for their time. All tests were performed in the same room to ensure environmental consistency across sessions.

4.1.1 Recruitment

We recruited participants through mailing lists at our institution. Pilot testing with participants showed that the third-party software used to recognize the activation keyword (and activate the device) was not robust enough to detect non-native accents saying "Alexa." Thus, we decided to recruit only English native speakers to reduce confounding factors based on accent and pronunciation, given that our goal was to assess the relative impact of dynamic privacy controls on activation keyword detection, rather than evaluating the robustness of the prototype's speech recognition software.

Ten participants were recruited for this lab study. Six of them were female, four were male. Their ages ranged from 18 to 26 years; all were students. While this constitutes a skewed sample, it did not interfere with this experiment's purpose of assessing whether there were relative differences in activation performance between a smart speaker with our novel privacy controls and without, in order to test the approaches' general feasibility.

4.2 Results

Overall, the activation success rates of Control, Voice and Gaze conditions are quite similar to each other. This means our privacy controls did not seem to interfere with participants' ability to successfully activate the smart speaker. However, participants found it easier to activate the Echo Dot than our prototype conditions, including the prototype-based control condition, likely due to better microphones and activation keyword recognition software in the Echo Dot.

In 57% of the measurements, participants managed to activate the device within one attempt; and within three attempts for 88% of the measurements. Figure 4 shows how the number of attempts differ across distances. The farther participants are away from the device, the more attempts it requires to activate the device. At 1.5 meters, the median number of attempts is 1 for all conditions. At 3 meters, the median is still 1 for most conditions, except for Gaze (M=2). At 4.5 meters, the median for Echo Dot and Voice remains at 1 attempt and is 2 attempts for Gaze and Control.

We built linear mixed-effect regression models to further examine the impact of distance on activation success rate. Using 1.5m as the reference point, the number of attempts increases as participants move to 3m

(β=.29, p=.06), and further increases at 4.5m significantly (β=.74, p<.001).

We next describe the performance differences between conditions across *all* distances. For Gaze, we selected data points at 0° angle and forward body orientation for the cross-condition comparison, since participants used the same angle and body orientation for the other conditions. Using Echo Dot as the reference point, Voice (β=−.66, p<.001), Gaze (β=.97, p<.001), and Control (β=.69, p=.001) all required significantly more attempts to activate the device. This suggests the commercial smart speaker has better speech recognition technology than our prototypes. When setting the prototype-based Control as the reference point instead, no significant differences in activation attempts between Control and Voice (β=−.02, p=.89), or Control and Gaze (β=.28, p=.11), were detected, i.e., we find no significant negative impact of the privacy controls on activation performance in our study. Furthermore, setting Voice as the reference point and comparing it with Gaze, the activation attempts in these two conditions do not differ from each other significantly either (β=.30, p=.08).

Within the Gaze condition, we also assessed the number of activation attempts at different angles (-20°, 0°, +20°) and head orientations (left, forward, right) for each distance in order to assess their impacts on gaze detection. For angle assessment, using 0° as the reference point, no significant differences in activation attempts between 0° and -20° (β=.14, p=.31), or 0° and 20° (β=.15, p=.29), were detected. For body orientation assessment, using forward as the reference point, participants facing left required significantly more activation attempts (β=.31, p<.05); no significant differences in activation attempts between forward and right were detected (β=.17, p=.24).

Because our sample size is relatively small, one could hypothesize that a larger sample might lead to statistically significant differences between Voice, Gaze and our prototype's control condition. However, the intraclass correlation coefficient (ICC) [51] for the random effect variable is .97, which means between individual participants, their performances across different distances and conditions followed a fairly consistent pattern. We do not assert that a larger sample will yield the exact same results, but we expect that the findings will not be much different.

Although the Echo Dot performed better overall than our privacy controls, this is most likely due to its superior microphones and audio processing. Among the conditions based on our prototype smart speaker

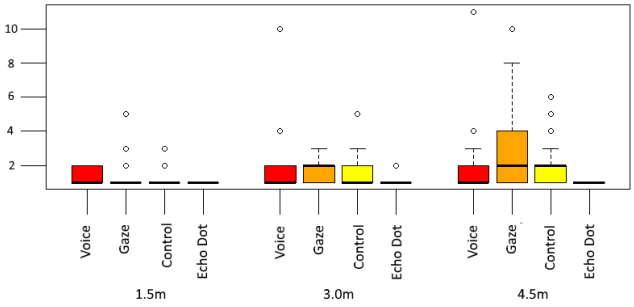

Fig. 4. Number of attempts to successfully activate the smart speaker at different distances by condition.

(Voice, Gaze, Control), the number of activation attempts were very similar at the 1.5m and 3m distances. Based on these findings, our new privacy controls do not seem to prevent users from successfully activating the smart speaker. This suggests that leveraging interpersonal cues for smart speaker privacy controls is feasible in practice.

5 Study: Usability and Perceived Privacy

We conducted a second lab study to examine (1) how our privacy controls perform in interactive scenarios; (2) how well participants could use our privacy controls; and (3) how participants perceived the level of privacy afforded by our privacy controls. We compared our proposed privacy controls (Voice and Gaze) to a standard mute button (Control). We wanted to learn how effectively participants could use these privacy controls to safeguard their privacy in contexts where they would need to interact with a smart speaker *and* have sensitive conversations at the same time.

We conducted a within-subjects study where pairs of participants completed a series of tasks that required them to both interact with the smart speaker as well as talk to each other without the smart speaker listening. Most participants preferred to use Voice over Gaze, with many participants finding the Voice condition easy to use. For perceived privacy levels, participants ranked Control the highest due to its mute button, and Voice was ranked higher than Gaze. Despite this perception, many participants remarked that they might forget to press the mute button. During the study, many participants indeed failed to use the Control's mute button when required, showing that our privacy controls that dynamically mute and unmute the smart speaker are

more effective as they are not susceptible to task completion errors.

5.1 Study Design

We tested three conditions in this second study: Voice, Gaze, and Control (our prototype with a physical mute button). All conditions used local activation keyword detection with the same keyword ("Alexa"). Since we had already tested the Echo Dot's performance in comparison to our prototypes (it performed better due to better hardware/software) and given that it has the same privacy mechanism as the Control condition, we did not include the Echo Dot in this study.

5.1.1 Study Protocol

Two participants were randomly paired to complete a study session together. Participant pairs were seated at a table with our smart speaker system between them. They completed one task per condition. The order of the three conditions and tasks was counter-balanced across participant pairs (Latin square).

Per condition, participant pairs were first introduced to the privacy control in the condition and familiarized themselves with it. Next, participants were given one of three tasks to perform (select a restaurant and reserve a table; plan a trip to a European city; decide on a movie and a showing time). These tasks were designed to include segments that had participants interact with the smart speaker, and segments in which they were *instructed* to discuss something among themselves and keep it private from the smart speaker. For example, in the restaurant task, participants would first ask the smart speaker for nearby restaurants; then, they privately discussed their personal preferences and chose a restaurant; next, they would ask the smart speaker for directions to the restaurant and to reserve a table. The full task descriptions given to participants are provided in Appendix A. While artificial, these tasks were designed to mimic situations in which individuals have private conversations while also interacting with the smart speaker, e.g., playing and controlling music via the smart speaker while in parallel having a private conversation among family members.

After completing a task, participants were asked in a post-task interview to rate their experience with the condition's privacy control using the System Usability Scale (SUS) [8] and answer open-ended questions regarding their experience (see Appendix B). SUS is a reliable and validated scale for measuring perceived usability and comparing it across systems [39], because it is technology agnostic, it is well suited for evaluating novel technologies such as ours [6]. The SUS scores, alongside the open-ended questions, allowed us to both qualitatively and quantitatively assess the perceived usability of the respective privacy control. Once participants had completed a task with each of the three conditions and the respective post-task questions, participants completed a card sorting exercise as part of the exit interview, in which they ranked the three conditions in terms of usability and privacy (see Appendix C). For each ranking, participants were asked to explain their rationale. This provides additional insights on how participants perceived the usability and the privacy level of the different controls.

All sessions were video-recorded to aid subsequent analysis. We also logged the smart speaker's microphone recordings and analyzed whether the smart speaker's microphone was correctly muted during the 'private' segments.

5.1.2 Recruitment

We recruited 18 English native speakers who had not participated in Study 1 as participants (15 female, 3 male) via mailing lists at our institution. Their ages ranged from 20 to 30 years. Participants were compensated $15 for their time. Sessions lasted 43–57 minutes (median: 46 min.). While college students may exhibit high technology affinity, our goal with this study was to understand how the same individual would perceive and rank the usability and privacy of the different privacy controls as an initial determination of the utility of using interpersonal cues as privacy controls.

5.2 Results

The Voice and Control conditions were rated as equally usable. Most participants preferred the Voice condition over the Control's mute button because they found Voice more natural and easy to use. Gaze's usability was rated lower due to difficulties in determining when the microphone was muted or unmuted.

For perceived privacy, participants found the physicality and immediate feedback of the mute button reassuring; the Voice condition was also perceived as providing privacy, but was ranked below the mute button; the

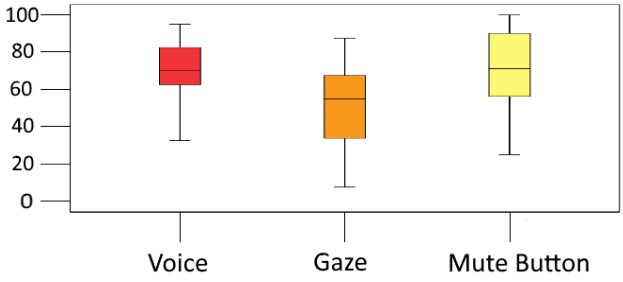

Fig. 5. System Usability Scale (SUS) score distributions for the different conditions.

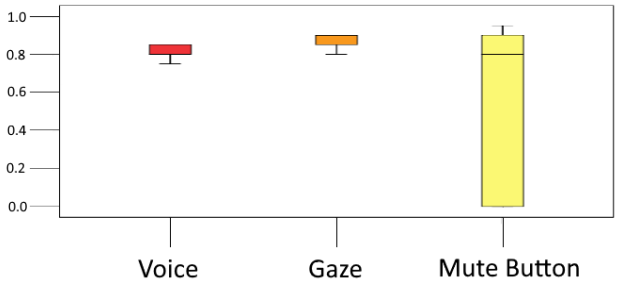

Fig. 6. Success rate for keeping conversations private across different conditions.

camera in the Gaze condition raised privacy concerns. Nevertheless, participants were more successful in keeping conversation parts private from the smart speaker with the Gaze condition than the Control: almost half the participants in the Control condition forgot to use the mute button. Although the difference between Voice and Control was not significant, with Voice, fewer participants made errors.

5.2.1 Usability

Figure 5 shows the conditions' SUS score distributions. A one-way repeated measures ANOVA showed significant differences between SUS scores ($F(2, 30)$=7.73, p<.05, η_p^2=.34). Post-hoc pairwise comparisons (Tukey HSD; Bonferroni corrected) showed that the usability of Voice (mean SUS=70.5; HSD p<.05) and the Control's mute button (mean SUS=71.6; HSD p<.05) were rated significantly higher than Gaze (mean SUS=50.8).

There was a general consensus that the Voice condition was useful and natural to use. One participant said *"I find it easier to use than a mute button, feels very natural to me."* Another stated, *"I have an Echo device and I don't use the button at all. This is so useful!"* Some participants suggested that the voice volume threshold should be adjustable or adjust automatically to different speakers: *"I tend to talk loudly. It would be great if it adapts to my volume levels;"* and *"it can be more streamlined if I can tune the audio level."*

Many participants liked the Gaze control, but some were concerned that it might limit when they can use the smart speaker: *"It is nice but can be cumbersome, I can look somewhere else."* and *"Often times I am working on something and I don't want to always look at 'her' every time."* One participant said: *"I can't really say if the device remains red when I look away."* This denotes a difficulty with perceiving feedback about the microphone state in the Gaze condition, as the state changes when looking at the device. One participant was confused about the optimal distance for the gaze detection to work: *"I don't know how far away I should be from the device."*

Some participants found the Control's mute button less convenient to use than the other controls: *"It is useful, but having to sit by it is very cumbersome"* and *"Cumbersome and have to walk up to it, never used the button in my Google Home."* One participant said they *"wouldn't mute the button, would not even think about the mute button,"* underlining that the mute button is easily forgotten and not integrated well into people's interaction with smart speakers [37].

Regarding usability, 12 of 18 participants ranked the Voice condition highest; and 4 ranked it second highest. The Control's mute button was ranked highest by 5 participants; second highest by 4. Only one participant ranked the Gaze condition highest; 10 ranked it second. The most common ranking of the three conditions (8 times) was Voice, Gaze, and Control. The strong preference for the Voice condition suggests that using voice volume to control privacy is perceived as intuitive, likely due to the appropriation of a very familiar interpersonal communication cue and its integration into the voice-oriented smart speaker interaction. Gaze, on the other hand, was rarely ranked first, because having to look at the device was perceived as limiting and state feedback was difficult to see (i.e., whether the device was muted when looking away). The physical mute button was perceived as cumbersome by many participants.

This study further confirmed that the dynamic privacy controls have little to no impact on smart speaker activation performance. No participants voiced issues regarding their ability to activate the smart speaker when desired.

5.2.2 Ability to Keep Conversations Private

Figure 6 shows the distributions of success rate in keeping the conversations private across conditions. Here, success rate is defined as the ratio of the instances the microphone was muted during a private conversation segment to the total number of private conversation segments. The median success rates were 90% for Gaze, 80% for Voice, and 80% for the Control's mute button. Notably, for the Control condition, in 4 out of 9 sessions the success rate was 0%, meaning that participants forgot to press the mute button during the session.

A repeated measures one-way ANOVA with Greenhouse-Geisser correction shows significant differences between conditions ($F(1.01, 8.10)=5.69$, $p<.05$, $\eta_p^2=.42$). Tukey HSD post-hoc tests with Bonferroni correction reveal a significant difference between Gaze and Control ($p<.05$).

Thus, even though Gaze has lower perceived usability, it was more effective at helping participants keep their conversation private than the mute button. Although the differences between the Voice and Control conditions were not significant, the distributions in Figure 6 show that with a physical mute button, there is a risk that users forget to press it, whereas fewer errors occurred with the Gaze and Voice conditions.

All conditions had similar false positive (smart speaker wakening when the participants did not desire to activate it) and false negative rates (smart speaker not waking up when the participants wanted to activate it). The median false positive rates were 0 for Voice (values ranged from 0 to 1), 0 for Gaze (values ranged 0 to 1), and 0 for Control (the only value was 0). This shows that accidental awakening of the smart speaker was not an issue for participants in our study. This is likely because in the private conversations there was little reason to use the activation keyword ("Alexa") or similar sounding phrases, so even if the microphone was activated, the keyword would not be said and so the smart speaker would not wake up.

The median false negative rates were 1 for Voice (values ranged from 0 to 6), 1.5 for Gaze (values ranged from 0 to 7), and 0.5 for Control (values ranged from 0 to 10). This suggests that participants occasionally struggled to activate the smart speaker. Given that false negative rates were fairly similar across conditions, and our results from the previous study, false negatives are likely due to the prototype's voice recognition capabilities rather than the privacy controls.

5.2.3 Perceived Level of Privacy

Many participants perceived the Voice condition as providing a high level of privacy. One participant said *"I like that I don't have to always look at the device and control my privacy."* Another participant said that it is usable but has a learning curve: *"It makes me feel private, however it needs a little getting used to to understand the volume threshold."* One participant preferred this control over the other two: *"better than having a camera look at me or press a button."*

For Gaze, in alignment with prior work [38, 53], many participants were concerned about the camera. One participant said *"Now someone is always listening and seeing me."* Another participant said *"I am more concerned now, I can get it to trigger without me knowing and the camera is always on and watching me."* One participant noted that the camera sensor affects potential placement: *"I like it, can be really useful but it depends on where I place it."* Due to our study design, participants directly compared the camera-based control to non-camera based controls. Future work needs to explore how a camera-based privacy control is perceived in the context of a camera-equipped smart speaker (e.g., Amazon's Echo Look) [38, 53].

For the Control, many participants indicated that they were very comfortable using a physical mute button, since they felt they have control over it: *"overall yeah it works. The fact that I can control it makes it more private."* Participants found the associated state feedback to be reassuring: *"I like that I can control it and be sure it's off through the color"* and *"When I press the button and it turns red, I have a feeling of control."*

Overall, when asked to rank the conditions by perceived level of privacy, 12 participants ranked the Control's mute button first, 4 ranked it second. Six participants ranked Voice first, 7 ranked it second. No participant ranked Gaze first, but 7 ranked it second. The most common ranking (7 times) was Control, Voice, Gaze. One reason for this order is that many participants found the immediate feedback associated with the physical button reassuring. Yet, our findings show that many participants forgot to press the physical mute button when they were supposed to keep information private from the smart speaker, a problem that did not occur with the two dynamic privacy controls.

6 Discussion

Our findings show that using interpersonal communication cues as privacy controls is a viable approach – they integrate well into users' interactions with smart speakers, are intuitive to use, and enhance user privacy. Based on our findings, we discuss challenges and practical considerations for designing such dynamic privacy controls, after describing potential limitations of our work.

6.1 Limitations

Our main contribution is exploring the concept of leveraging interpersonal communication cues to develop privacy controls, and validating specific aspects of this novel approach (impact on activation performance, usability, perceived privacy). We deliberately chose lab studies to achieve high internal validity and examine effects of our new privacy controls under controlled conditions. While we took measures to increase external validity, such as explicitly instructing participants to keep certain parts of conversations private, we acknowledge that the artificial lab setting and relatively simple tasks do not yield findings that are generalizable to all real-world smart speaker use cases. Despite this, our findings still demonstrate the general feasibility and utility of the proposed privacy control approach, laying the groundwork for future studies with such privacy controls in different contexts and scenarios.

Related to this, because our participants were all native English speakers and U.S. college students, we refrain from making claims about the generalizability of our findings to other sub-populations or the general population. Future research may validate our findings with different populations using larger samples. However, we observed that in the performance study participants across conditions exhibited a consistent pattern, evidenced by high ICC (.97). Therefore, we believe our quantitative analysis, even though exploratory rather than confirmatory in nature, is still rigorous, and we do not expect a larger sample will lead to significantly different findings.

Regarding the student sample, students may be more likely to adopt or be familar with smart speakers due to potentially higher technology savviness and digital literacy. However, there is little reason to believe that students are more or less likely to personify smart speakers than the general population, especially given that personification of smart speakers has also been found in non-student populations [56]. We do not expect other population characteristics to substantially alter our findings. Nevertheless, further research is needed to better understand whether and how our approach needs to be adapted to serve more diverse populations and contexts, including non-native English speakers, other languages, and other sociocultural contexts.

Our prototype poses an additional limitation. In particular, the Kinect camera was not robust enough to accurately track eye movement and detect gaze direction. However, our results show that head orientation served as a reasonable proxy for gaze for our purposes, given that people often naturally align their head orientation with their gaze. Future studies could explore more accurate cameras that track gaze direction.

Another possible limitation has to do with the instructions we gave participants to keep conversations quiet. Given the explicitness of the instructions, participants' behavior might have differed from circumstances in daily life, i.e., in real world circumstances people might not have the need for privacy at the forefront of their minds, and so the privacy controls might not be as effective. We opted to provide explicit instructions explicit in order to test whether the privacy controls worked when users wanted to use them – a necessary first step in evaluating how effective these controls are. Future work could examine participants' real world behaviors through field studies to confirm the extent to which the proposed privacy controls are effective at muting the microphone when it is not needed under different conditions.

6.2 Interpersonal Cues as Privacy Controls

Any privacy control implemented in a device requires some trust from users. Currently, smart speaker users need to trust that a smart speaker only listens for the activation word. With our proposed controls, users need to trust that the sensor which activates the microphone only does so when the interpersonal cue has been detected before the wake word can even be detected. While the level of trust needed from users remains the same, our controls do provide privacy benefits: mitigating 'always-listening' concerns and false activation by adding a secondary privacy control that shifts device activation out of the primary interaction modality (voice interaction).

Our prototype further demonstrates the promise of this idea — using interpersonal communication cues as

privacy controls is feasible and practical. Voice and gaze privacy controls had a negligible effect on smart speaker functionality. Furthermore, participants took to these controls intuitively and easily (especially the Voice control), with some participants finding our controls more convenient to use than physical mute buttons. Based on these findings, our controls demonstrate the potential to help users keep a smart speaker's microphones effectively muted when they are not interacting with the device. Given the ease and convenience of our privacy controls, users may be more likely to activate such privacy controls for their smart speaker. This is a substantial privacy improvement over current smart speakers, which are continuously listening for their activation keyword regardless of whether they are needed, and the privacy controls of which (i.e., the mute button) are rarely utilized by general users [37], posing unnecessary risks of surreptitious recording as well as false activation.

Although our study focused on two specific interpersonal cues for a specific IoT device (smart speakers), this approach may also be relevant for building privacy controls — possibly based on different interpersonal communication cues — for other IoT devices. The same broad principles we leveraged here (users are familiar with interpersonal communication cues, are amenable to exhibiting them; cues can be leveraged to detect when a user wants to interact with a device) may apply to other cues and devices. For example, one could imagine a smart toy that can record video and audio[7] leveraging interpersonal cues (such as a child's gestures) to detect when it is being played with, and automatically cease audio/video recording when it is not being played with.

6.3 Cues Must Match Context

When developing novel privacy controls for IoT devices, it is important to keep the device's context of use in mind. Although participants used our controls well to maintain their privacy, situations can be imagined in which our proposed mechanisms would not work as well. For example, with voice-volume-based privacy controls, the microphone could be inadvertently activated by loud ambient noise (such as a TV), or if two people are engaged in a loud argument. While possible, this would not automatically result in audio recording, but rather activate listening for the activation keyword, which corresponds to the status quo of smart speakers. As a result, the voice-volume-based control still offers a privacy improvement over current smart speakers as speech recognition would be deactivated for most of the time. Similarly, gaze-based privacy controls only work when users are in the same room as the smart speaker and have a line of sight to the device.

Thus, privacy controls that use interpersonal communication cues need to consider the setting the device will be in, as well as the sociocultural context and specific cues of people using the device. One approach for accomplishing this is to allow users to customize their privacy control configurations, and to have privacy mechanisms that support multiple cues. For instance, with a voice-based privacy control, the minimum threshold needed to activate the smart speaker could be set by users, or change automatically based on background ambience level. Vision-based privacy controls could be combined with other non-vision based interpersonal cues, so that users can interact with a smart speaker regardless of whether they are in the same room.

Such dynamic privacy controls can further be combined with other privacy mechanisms into a layered privacy protection approach. For instance, our voice-volume-based and gaze-based privacy controls each function well in combination with the device requiring an activation keyword, yet help to limit and reduce false activation or surreptitious data collection when the device is unlikely to be used, in contrast to current smart speakers which are always listening for activation keywords and are prone to false activation.

6.4 Need for Privacy Feedback

While participants found the mute button less convenient than the new privacy controls, participants liked the tangible and immediate feedback it provided. The consistent state (microphone on or off until button pressed again) made it easy to determine when the microphone was muted or unmuted. While our smart speaker's visual feedback was modeled after existing commercial smart speakers, the adaptive nature of the privacy controls creates uncertainty for some participants about when the microphone is on or off. For the gaze-based privacy control in particular, checking whether the microphone is on or off was inherently difficult because looking at the device activates the microphone.

Thus, a further research challenge is how to better convey to users the current state of privacy settings and when they have changed, in ways that are suitable for

7 Such as the ones sold by https://www.smarttoy.com/

the specific interpersonal cue used and the user's context. A potentially promising approach is to use ambient interfaces [29] that indicate to users when data is being collected. For example, a smart camera could use ambient lighting, or emit background white noise, when it is gathering data, and turn off the ambient lighting or be silent when it is no longer gathering data.

6.5 Additional Sensors Raise Concerns

A challenge when designing context-adaptive privacy systems is that they leverage sensors to detect content, and sensors themselves raise privacy concerns [59, 61]. With our gaze-based privacy control many participants were concerned about the use of a camera, finding it to be more invasive than a smart speaker's microphone, which we expected based on prior work [38, 53]. In contrast, there were less concerns over our voice-volume-based privacy control, likely due to (a) the voice-based privacy control not requiring additional sensors beyond a microphone (which smart speakers already have), and (b) the data gathered by the voice-based privacy control (voice volume level) being less sensitive than a camera feed. Notably, certain smart speakers and other smart home devices already have cameras integrated, and camera-based privacy controls might be perceived differently in the context of such devices.

When additional sensors are needed to detect interpersonal communication cues or to dynamically adapt privacy settings, it is important to ensure that data collection is minimal and that the collected data cannot be compromised. This could be accomplished by local processing of data, ensuring that raw sensor streams are isolated from network access to reduce potential misuse, and by making code of privacy control components open source to allow inspection and instill confidence that the privacy protection works as advertised. Another approach could be the use of low fidelity sensors that are limited in their data-capturing capabilities, to determine when higher fidelity sensors should be active. For our voice-based privacy mechanism, we utilized a secondary microphone that only detected loudness levels to determine when the smart speaker's primary microphone and speech recognition should be activated. Using a decibel meter instead, which measures loudness and cannot record audio, would further reduce potential for unintended data collection.

7 Conclusion

We propose a novel approach for controlling privacy around smart speakers: leveraging interpersonal communication cues to dynamically determine when a smart speaker's microphone needs to be active and listen for its activation keyword. We implemented and tested the feasibility and usability of voice-volume-based and gaze-based privacy controls. Both approaches have only negligible impact on the activation performance of a smart speaker. Additionally, participants found our privacy controls more easy to use and intuitive than existing controls (mute button), demonstrating the potential of these dynamic privacy controls to better safeguard user privacy and limit false activation.

Our findings provide a promising direction for designing privacy controls for smart speakers and IoT devices that complement and support other privacy mechanisms, such as requiring an activation keyword for voice commands. Further research needs to investigate the feasibility, utility and usability of privacy controls leveraging interpersonal communication cues in different contexts, as well as opportunities for leveraging other interpersonal cues beyond voice volume and gaze direction.

Acknowledgments

This research has been partially funded by the University of Michigan School of Information and by the National Science Foundation under grant CNS-1330596. The authors thank the reviewers for their constructive feedback, as well as the members of the Security Privacy Interaction Lab (spilab) for their support.

References

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

[2] Mark Altosaar, Roel Vertegaal, Changuk Sohn, and Daniel Cheng. Auraorb: Using social awareness cues in the design of progressive notification appliances. In *Proceedings of the 18th Australia Conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments*, OZCHI '06, pages 159–166, New York, NY, USA, 2006. ACM.

[3] Tawfiq Ammari, Jofish Kaye, Janice Y. Tsai, and Frank Bentley. Music, search, and iot: How people (really) use voice assistants. *ACM Trans. Comput.-Hum. Interact.*, 26(3):17:1–17:28, April 2019.

[4] Newgenn Apps. 13 IoT Statistics Defining the Future of Internet of Things. Newgenn Apps, https://www.newgenapps.com/blog/iot-statistics-internet-of-things-future-research-data, 2018. Online; accessed 09/01/2018.

[5] Richard Baguely and Colin McDonald. "Appliance Science: Alexa, how does Alexa work? The science of the Amazon Echo". CNET, https://www.cnet.com/news/appliance-science-alexa-how-does-alexa-work-the-science-of-amazons-echo/, 2016. Online; accessed 02/26/2019.

[6] Aaron Bangor, Philip Kortum, and James Miller. Determining what individual sus scores mean: Adding an adjective rating scale. *J. Usability Studies*, 4(3):114–123, May 2009.

[7] Brian Barret. What Amazon Echo and Google Home do with your voice data. Wired, https://www.wired.com/story/amazon-echo-and-google-home-voice-data-delete/, November 2017. Online: accessed 05/08/2018.

[8] John Brooke. SUS – a quick and dirty usability scale. In *Usability evaluation in industry*. Taylor & Francis, London, 1996.

[9] Joseph Bugeja, Andreas Jacobsson, and Paul Davidsson. On privacy and security challenges in smart connected homes. In *Intelligence and Security Informatics Conference (EISIC)*, pages 172–175. IEEE, 2016.

[10] Daniel J Butler, Justin Huang, Franziska Roesner, and Maya Cakmak. The privacy-utility tradeoff for remotely teleoperated robots. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 27–34. ACM, 2015.

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

[12] Goodwin Charles. *Conversational organization: Interaction between speakers and hearers*. New York, Academic Press, 1981.

[13] Peng Cheng, Ibrahim Ethem Bagci, Jeff Yan, and Utz Roedig. Smart speaker privacy control-acoustic tagging for personal voice assistants. In *IEEE Workshop on the Internet of Safe Things (SafeThings 2019)*, 2019.

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

[15] Alice Coucke, Alaa Saade, Adrien Ball, Théodore Bluche, Alexandre Caulier, David Leroy, Clément Doumouro, Thibault Gisselbrecht, Francesco Caltagirone, Thibaut Lavril, et al. Snips voice platform: an embedded spoken language understanding system for private-by-design voice interfaces. *arXiv preprint, arXiv:1805.10190*, 2018.

[16] Ry Crist and Andrew Gebhart. Everything you need to know about the Amazon Echo. CNET, https://www.cnet.com/how-to/amazon-echo-alexa-everything-you-need-to-know/, 2017. Online; accessed 02/26/2019.

[17] Mark L Daly and Knapp John Augustine. *Handbook of interpersonal communication*. Sage, 2002.

[18] Tamara Denning, Cynthia Matuszek, Karl Koscher, Joshua R Smith, and Tadayoshi Kohno. A spotlight on security and privacy risks with future household robots: attacks and lessons. In *Proceedings of the 11th international conference on Ubiquitous computing*, pages 105–114. ACM, 2009.

[19] Joseph A DeVito. *Interpersonal communication*. New York: Longman Inc, 2007.

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

[21] Jide S. Edu, Jose M. Such, and Guillermo Suarez-Tangil. Smart home personal assistants: A security and privacy review. arXiv preprint, arXiv:1903.05593, 2019.

[22] Serge Egelman, Raghudeep Kannavara, and Richard Chow. Is this thing on?: Crowdsourcing privacy indicators for ubiquitous sensing platforms. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1669–1678. ACM, 2015.

[23] Stefano Fasciani. *Voice Controlled interface for Digital Musical Instrument*. PhD thesis, 2014. Online; accessed 05/09/2018.

[24] Brian Feldman. "like everyone else, amazon's alexa is laughing at you". New York Magazine, http://nymag.com/intelligencer/2018/03/amazon-alexa-is-laughing-at-you.html, 2018. Online; accessed 02/26/2019.

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

[26] Catherine S Fichten, Vicki Tagalakis, Darlene Judd, John Wright, and Rhonda Amsel. Verbal and nonverbal communication cues in daily conversations and dating. *The Journal of Social Psychology*, 132(6):751–769, 1992.

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

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

[29] Tom Gross. Ambient interfaces in a web-based theatre of work. In *Proceedings of the 10th Euromicro Workshop on Parallel, Distributed and Network-based Processing*, pages 55–62. IEEE, 2002.

[30] Susumu Harada, Jacob O Wobbrock, and James A Landay. Voice games: investigation into the use of non-speech voice input for making computer games more accessible. In *IFIP Conference on Human-Computer Interaction*, pages 11–29. Springer, 2011.

[31] Richard Jones. *Communication in the real world: An introduction to communication studies*. The Saylor Foundation, 2013.

[32] Bjorn Karmann. Project alias. https://bjoernkarmann.dk/project_alias, 2018. Online; accessed 08/22/2019.

[33] Jacob Kastrenakes. Amazon now lets you tell Alexa to delete your voice recordings. The Verge, https://www.theverge.com/2019/5/29/18644027/amazon-alexa-delete-voice-recordings-command-privacy-hub, 2019. Online; accessed 08/28/2019.

[34] Bastian Könings, Florian Schaub, and Michael Weber. Privacy and trust in ambient intelligent environments. In *Next Generation Intelligent Environments*, pages 133–164. Springer, 2016.

[35] Saadi Lahlou, Marc Langheinrich, and Carsten Röcker. Privacy and trust issues with invisible computers. *Commun. ACM*, 48(3):59–60, March 2005.

[36] Marc Langheinrich and Florian Schaub. Privacy in mobile and pervasive computing. *Synthesis Lectures on Mobile and Pervasive Computing*, 10(1):1–139, 2018.

[37] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. Alexa, are you listening?: Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):102, 2018.

[38] Hosub Lee and Alfred Kobsa. Privacy preference modeling and prediction in a simulated campuswide iot environment. In *2017 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 276–285. IEEE, 2017.

[39] James R. Lewis. The system usability scale: Past, present, and future. *International Journal of Human–Computer Interaction*, 34(7):577–590, 2018.

[40] Yuting Liao, Jessica Vitak, Priya Kumar, Michael Zimmer, and Katherine Kritikos. Understanding the role of privacy and trust in intelligent personal assistant adoption. In *International Conference on Information*, pages 102–113. Springer, 2019.

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

[42] Aadil Mamuji, Roel Vertegaal, J Shell, Thanh Pham, and Changuk Sohn. Auralamp: Contextual speech recognition in an eye contact sensing light appliance. In *Extended abstracts of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, 2003.

[43] Zied Mani and Ines Chouk. Drivers of consumers' resistance to smart products. *Journal of Marketing Management*, 33(1-2):76–97, 2017.

[44] Martin, Taylor. 12 ways to use Alexa in the kitchen. CNET, https://www.cnet.com/how-to/amazon-echo-ways-to-use-alexa-in-the-kitchen/, November 2017. Online; accessed 05/05/2018.

[45] Evelyn McClave, Helen Kim, Rita Tamer, and Milo Mileff. Head movements in the context of speech in arabic, bulgarian, korean, and african-american vernacular english. *Gesture*, 7(3):343–390, 2007.

[46] Emily McReynolds, Sarah Hubbard, Timothy Lau, Aditya Saraf, Maya Cakmak, and Franziska Roesner. Toys that listen: A study of parents, children, and internet-connected toys. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 5197–5207, New York, NY, USA, 2017. ACM.

[47] Pardis Emami Naeini, Sruti Bhagavatula, Hana Habib, Martin Degeling, Lujo Bauer, Lorrie Cranor, and Norman Sadeh. Privacy expectations and preferences in an iot world. In *Symposium on Usable Privacy and Security (SOUPS)*, 2017.

[48] Kayako Nakagawa, Masahiro Shiomi, Kazuhiko Shinozawa, Reo Matsumura, Hiroshi Ishiguro, and Norihiro Hagita. Effect of robot's whispering behavior on people's motivation. *International Journal of Social Robotics*, 5(1):5–16, Jan 2013.

[49] Michael Natale. Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology*, 32(5):790, 1975.

[50] David H Nguyen and Elizabeth D Mynatt. Privacy mirrors: understanding and shaping socio-technical ubiquitous computing systems. Technical report, Georgia Institute of Technology, 2002.

[51] Carol AE Nickerson. A note on "a concordance correlation coefficient to evaluate reproducibility". *Biometrics*, pages 1503–1507, 1997.

[52] Katarzyna Olejnik, Italo Dacosta, Joana Soares Machado, Kévin Huguenin, Mohammad Emtiyaz Khan, and Jean-Pierre Hubaux. Smarper: Context-aware and automatic runtime-permissions for mobile devices. In *IEEE Symposium on Security and Privacy*, pages 1058–1076. IEEE, 2017.

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

[54] Richard A. Page and Joseph L. Balloun. The effect of voice volume on the perception of personality. *The Journal of Social Psychology*, 105(1):65–72, 1978.

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

[56] Amanda Purington, Jessie G. Taft, Shruti Sannon, Natalya N. Bazarova, and Samuel Hardman Taylor. "alexa is my new bff": Social roles, user satisfaction, and personification of the amazon echo. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2853–2859, New York, NY, USA, 2017. ACM.

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

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

[59] Florian Schaub. Context-adaptive privacy mechanisms. In Aris Gkoulalas-Divanis and Claudio Bettini, editors, *Handbook of Mobile Data Privacy*, pages 337–372. Springer International Publishing, Cham, 2018.

[60] Florian Schaub, Rebecca Balebako, Adam L Durity, and Lorrie Faith Cranor. A design space for effective privacy notices. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*, pages 1–17, 2015.

[61] Florian Schaub, Bastian Könings, and Michael Weber. Context-adaptive privacy: Leveraging context awareness to support privacy decision making. *IEEE Pervasive Computing*,

14(1):34–43, Jan 2015.

[62] Florian Schaub, Bastian Könings, Peter Lang, Björn Wiedersheim, Christian Winkler, and Michael Weber. Prical: context-adaptive privacy in ambient calendar displays. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 499–510. ACM, 2014.

[63] Bill Schilit, Norman Adams, and Roy Want. Context-aware computing applications. In *First Workshop on Mobile Computing Systems and Applications*, pages 85–90. IEEE, 1994.

[64] Albrecht Schmidt. Implicit human computer interaction through context. *Personal Technologies*, 4(2):191–199, Jun 2000.

[65] Singleton, Micah. Alexa can now set reminders for you. The Verge, https://www.theverge.com/circuitbreaker/2017/6/1/15724474/alexa-echo-amazon-reminders-named-timers, June 2017. Online; accessed 05/05/2018.

[66] Adam J. Sporka, Sri H. Kurniawan, and Pavel Slavík. Acoustic control of mouse pointer. *Universal Access in the Information Society*, 4(3):237–245, Mar 2006.

[67] Blase Ur, Jaeyeon Jung, and Stuart Schechter. Intruders versus intrusiveness: teens' and parents' perspectives on home-entryway surveillance. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 129–139. ACM, 2014.

[68] Roel Vertegaal, Aadil Mamuji, Changuk Sohn, and Daniel Cheng. Media eyepliances: Using eye tracking for remote control focus selection of appliances. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, pages 1861–1864, New York, NY, USA, 2005. ACM.

[69] Roel Vertegaal, Robert Slagter, Gerrit van der Veer, and Anton Nijholt. Eye gaze patterns in conversations: There is more to conversational agents than meets the eyes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 301–308, New York, NY, USA, 2001. ACM.

[70] Mark Weiser. Some computer science issues in ubiquitous computing. *Commun. ACM*, 36(7):75–84, July 1993.

[71] Joan Welkowitz, Stanley Feldstein, Mark Finklestein, and Lawrence Aylesworth. Changes in vocal intensity as a function of interspeaker influence. *Perceptual and Motor Skills*, 35(3):715–718, 1972.

[72] Primal Wijesekera, Arjun Baokar, Lynn Tsai, Joel Reardon, Serge Egelman, David Wagner, and Konstantin Beznosov. The feasibility of dynamically granted permissions: Aligning mobile privacy with user preferences. In *IEEE Symposium on Security and Privacy*, pages 1077–1093. IEEE, 2017.

[73] Danny Wyatt, Tanzeem Choudhury, and Henry Kautz. Capturing spontaneous conversation and social dynamics: A privacy-sensitive data collection effort. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 4, pages IV–213. IEEE, 2007.

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

[75] Serena Zheng, Marshini Chetty, and Nick Feamster. User perceptions of privacy in smart homes. *arXiv preprint, arXiv:1802.08182*, 2018.

A Usability Study Task Descriptions

A.1 Task A

Discuss with your partner and find nearby restaurants to go out for dinner.

1. Ask Alexa to show nearby Thai restaurants. You can have a look at the Alexa app on the smartphone for a comprehensive list with ratings from Yelp.
2. Discuss privately with your partner and decide on a restaurant. Consider including average rating, price, reputation, food allergies, distance into account for making a decision. The voice assistant should not learn your preferences, and you should keep your conversation private from the smart speaker's microphone.
3. Now ask Alexa for available tables by saying "Alexa, find me a table for 2 at 'Restaurant name' for dinner tonight."

A.2 Task B

You and your partner are planning a vacation in Europe for 15 days and would like to cover a minimum of five cities. Discuss with your partner and decide which city to visit first. The cities you are considering are Paris, Venice, Amsterdam, Madrid, and Rome.

1. You can start discussing privately with your partner now. Take into consideration your mutual interests in the particular city and consider backing up your decisions with reasons for choosing that city. You should keep your conversation private from the smart speaker's microphone.
2. Once you have decided on what city to visit first, ask Alexa a few questions about the location, local currency and weather in that city.
3. Now that you have decided which city to visit first, ask Alexa to plan a day in that city by saying, "Alexa open trip planner." Follow the voice instructions and request to plan a day at the selected city when prompted. The trip details can be seen on the Alexa app on the smartphone.

A.3 Task C

With the help of the voice assistant, discuss with your partner and decide on a movie to see this weekend at a nearby multiplex.

1. Ask Alexa a list of movies playing at the nearest movie theater for this weekend.
2. Privately discuss with your partner the movie to watch based on your preferences. Consider including preferred genre of movie, average rating for the movie, showtimes, etc. You should keep your conversation private from the smart speaker's microphone
3. Once decided, ask Alexa show timings for the selected movie this weekend.

B Usability Study Post-Task Interview Script

Participants were first given the 10-item System Usability Scale (SUS) to complete,[8] and then asked open ended questions:

1. Talk about your experience using the system, what do you think went well and what went wrong ?
2. Describe your emotions while interacting with the device, did you get frustrated or annoyed by the behavior of the device?
3. What was your experience using the privacy controls?
4. Do you trust the privacy controls to be effective / useful?

C Usability Study Exit Interview Script

Researcher: "Thank you so much for participating in our study! Before we go we have one last activity we want to you to do."

Researcher hands out 3 cards to each participant: each card has the name of one of the conditions, "Mute Button," "Speak Up," and "Look at Me."

R: "If you could take a few moments to rank the conditions in front of you based on usability – that is, how easy and comfortable you were using the device. And if you could do this individually please."

Wait for participants to complete task.

R: "Now if you could walk us through why you ordered the cards in that order?"

Discussion takes place.

R: "Thank you so much! Now if you could do the same thing, but this time, rank the cards in front of you in terms of privacy – that is, how confident that with the given control you will be able to maintain your privacy and keep unwanted conversations from Alexa."

Wait for participants to complete task.

R: "Now if you could walk us through why you ordered the cards in that order?"

Discussion takes place.

R: "Thank you so much!"

8 See https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html