Investigating Users' Preferences and Expectations for Always-Listening Voice Assistants

MADIHA TABASSUM, University of North Carolina at Charlotte, USA
TOMASZ KOSIŃSKI, Chalmers University of Technology, Sweden
ALISA FRIK, University of California, Berkeley / ICSI, USA
NATHAN MALKIN, University of California, Berkeley, USA
PRIMAL WIJESEKERA, University of California, Berkeley / ICSI, USA
SERGE EGELMAN, University of California, Berkeley / ICSI, USA
HEATHER RICHTER LIPFORD, University of North Carolina at Charlotte, USA

Many consumers now rely on different forms of voice assistants, both stand-alone devices and those built into smartphones. Currently, these systems react to specific wake-words, such as "Alexa," "Siri," or "Ok Google." However, with advancements in natural language processing, the next generation of voice assistants could instead always listen to the acoustic environment and proactively provide services and recommendations based on conversations without being explicitly invoked. We refer to such devices as "always listening voice assistants" and explore expectations around their potential use. In this paper, we report on a 178-participant survey investigating the potential services people anticipate from such a device and how they feel about sharing their data for these purposes. Our findings reveal that participants can anticipate a wide range of services pertaining to a conversation; however, most of the services are very similar to those that existing voice assistants currently provide with explicit commands. Participants are more likely to consent to share a conversation when they do not find it sensitive, they are comfortable with the service and find it beneficial, and when they already own a stand-alone voice assistant. Based on our findings we discuss the privacy challenges in designing an always-listening voice assistant.

CCS Concepts: • **Human-centered computing → Personal digital assistants**; *User studies.*

Additional Key Words and Phrases: voice assistants, always listening, survey

ACM Reference Format:
Madiha Tabassum, Tomasz Kosiński, Alisa Frik, Nathan Malkin, Primal Wijesekera, Serge Egelman, and Heather Richter Lipford. 2019. Investigating Users' Preferences and Expectations for Always-Listening Voice Assistants. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 4, Article 153 (December 2019), 23 pages. https://doi.org/10.1145/3369807

Authors' addresses: Madiha Tabassum, University of North Carolina at Charlotte, Charlotte, NC, USA, mtabassu@uncc.edu; Tomasz Kosiński, Chalmers University of Technology, Gothenburg, Sweden, tomasz.kosinski@chalmers.se; Alisa Frik, University of California, Berkeley / ICSI, Berkeley, CA, USA, afrik@icsi.berkeley.edu; Nathan Malkin, University of California, Berkeley, Berkeley, CA, USA, nmalkin@cs.berkeley.edu; Primal Wijesekera, University of California, Berkeley / ICSI, Berkeley, CA, USA, primal@berkeley.edu; Serge Egelman, University of California, Berkeley / ICSI, Berkeley, CA, USA, egelman@icsi.berkeley.edu; Heather Richter Lipford, University of North Carolina at Charlotte, Charlotte, NC, USA, richter@uncc.edu.

2474-9567/2019/12-ART153 $15.00
https://doi.org/10.1145/3369807

1 INTRODUCTION

Voice assistants, including stand-alone devices and those built into smartphones, are increasingly popular among consumers. Almost 41% of adults in the U.S. have access to a stand-alone "smart speaker" in their homes [34]. Currently, these systems respond when explicitly invoked by a specific wake-word—such as "Alexa," "Siri," or "Ok Google"—and serve common functions such as answering questions and controlling other connected devices.

Prominent players in the industry are trying to make voice assistants even more seamless. For instance, Google Home and Alexa offer a continued conversation feature (or "follow-up mode") to allow users to make follow-up requests after the first command without repeating the wake-word [12, 13]. Alexa also has a feature called Drop In, which allows whitelisted users to begin an audio or video call without the other party manually picking up the call [2], effectively allowing remote parties to listen at any time. Looking forward, Amazon and Google have patented the ability of voice assistants to automatically extract keywords from ambient speech and use that to provide targeted ads to users [39].

With advancements in natural language processing technology, we can therefore expect that the next generation of voice assistants will offer proactive assistance based on the audio signals and conversations acquired from the acoustic environment. In other words, they will work without being explicitly invoked by users [29]. For example, future voice assistants could recognize when a user is talking about dinner plans and may suggest updating the calendar, inviting friends, or making a reservation—or even doing all of these things automatically. Allowing devices to utilize the entirety of a conversation, including its context, will potentially enable fully-fledged assistants that can fluidly support users in their daily lives without detracting from their interactions with others.

Despite the potential to help users in their daily activities, always-listening voice assistants are virtually certain to raise privacy concerns. First, some consumers are resistant to adopting even the current keyword-triggered voice assistants, because they are uncomfortable with the devices' audio-data collection during regular use [10, 18] and in the background [32]. Continuous listening is likely to deepen and extend that concern since *all* conversations may be subject to recording, and because these recordings could be available to third parties. Second, compared with keyword-triggered listening, there is a higher chance of sensitive information being recorded without users' awareness. Users have expressed discomfort over controlling unprompted assistance [3] and unintended exposure of personal data [27] in such an environment.

Third, these devices are more likely to violate users' expectations of territorial privacy and embodied self-image. Territorial privacy represents the right of a person to determine individually when and how other entities are allowed to participate in his or her personal territory [16]. With keyword-triggered assistants, users actively control when the device is on. With devices that listen passively (i.e., without interrupting the conversation), users are likely to perceive less control over their private territory. With continuous access to audio data, it will be possible to identify and track people in their private spaces and make inferences about family dynamics, relationships, preferences, behaviors, and more.

Moreover, research has shown that people tend to allow access to their data when it is used for a goal that they find beneficial or when the data is not sensitive in general [30]. On the other hand, people are more likely to deny data collection if they think the data is very personal or sensitive [30]. In order to understand how always-listening devices may fit into users' day-to-day lives, it is important to investigate the trade-offs between the sensitivity of a conversation, the concerns about sharing its recording, and the benefits users receive. Specifically, it is critical to understand: what are people's privacy expectations in an always-listening environment? What factors affect those expectations and the decision to adopt this technology? What benefits and services do people expect in exchange for sharing their conversations? What concerns do they have regarding using such a device?

Our goal is to inform the design of always-listening systems that respect users' privacy expectations. In this paper, we therefore investigate the following research questions:

- What services do users expect always-listening voice assistants to provide based on their conversations?

- How useful are these potential services? Do people feel comfortable with them?
- How do factors such as demographics, perceived comfort, and service usefulness affect users' intentions to allow the voice assistant to access their conversations in exchange for services?

To answer these questions, we conducted an online survey, in which 178 participants listened to 2-3 minute conversation snippets and answered questions regarding the services they could envision an always-listening voice assistant providing to them.

Our results reveal that participants expect to receive a range of services, including those already available from existing voice assistants, as well as services unique to an always-listening device; for example, recognizing the context of the conversation and offering to add a calendar event, or making a purchase on the user's behalf. In general, participants perceived the services as useful. However, many participants, especially those who currently do not use existing smart speakers, were uncomfortable with the always-listening nature of the hypothetical voice assistants. Participants were more willing to allow voice assistant to use a conversation when they were comfortable with the services or found them beneficial and did not perceive the conversation as sensitive. Finally, we identify other concerns with always-listening assistants, pertaining to performance, service relevance, loss of agency, conversation interruptions, and security. We conclude with a discussion of implications, challenges, and recommendations for designing the next generation of assistant devices.

2 RELATED WORK

Several studies have investigated the role of intelligent personal assistants in human conversations. In a semi-structured interview study with 14 participants, Luger et al. [21] found that there is a disconnect between user expectations and abilities of current voice assistants: users expect the conversational agents to infer the context of the current task from the previous interactions. Porcheron et al. [36] investigated the use of voice assistants in multi-party conversations through an ethnomethodology study with 12 participants. They found that the assistant is primarily used for information search; however, actively making queries creates a lapse in the conversation even if only one participant is interacting with the device. Passive-listening devices may avoid such lapses in the conversation [23]. In another study, Porcheron et al. [35] found that voice assistants are embedded into everyday interactions in the home, but the interaction with the device is broken up among the other activities rather than continuous. These findings suggest potential benefits of passive-listening voice assistants that can provide services based on the surrounding context and conversation without the need for explicit queries.

Passive-listening technologies and services they may provide have been studied in various contexts such as supporting meetings, influencing conversations, providing guidance in various tasks, etc. For example, Kilgour et al. [15] proposed the Ambient Spotlight that uses meeting conversations to extract relevant documents from the computer. Shi et al. [38] developed IdeaWall, a system that provides visual cues derived from a background conversation in order to facilitate collaboration and creativity in a meeting. Carrascal et al. [8] envisioned a system that can transcribe phone calls and automatically parse fragments of text that the caller would be interested in reviewing later. They found a number of patterns that the callers would normally like to record for later, such as phone numbers, dates, addresses, prices, shopping/to-do lists, contact names, and activities. Through a combination of observation and interview studies, McGregor et al. [4, 24] explored a speech-based system that proactively detects actionable items and takes those actions in formal meetings. They found that contextual information is critical to accurately identify actionable items and correctly understand them after the meeting. In a "Wizard of Oz" study with 10 participants, Vtyurina et al. [42] found the importance of implicit conversational cues in achieving higher user satisfaction in the passive-listening environment.

While the services that can be provided by a passive-listening device might be very useful, continuous listening may raise privacy concerns and hinder the adoption of such a device. People are already concerned about existing smart speakers. Cowan et al. [10] conducted focus groups with 20 participants, who had tried, but chose not to use intelligent personal assistants on a regular basis. They identified lack of trust, data ownership, and data permanency as the potential barriers to using such technologies. In a more recent interview study with 17 users and non-users of smart voice assistants, Lau et al. [18] found that the potential for privacy violations is one of the main reasons behind not adopting the device. They also found that privacy controls available in these devices are not well aligned with users' needs and hence are rarely used. In a survey study with 116 smart speaker owners, Malkin et al. [22] found that users are not comfortable with the permanent retention of their recordings, especially those that include children and guests.

Always-listening agents will likely generate even greater concerns. In a laboratory study with 24 participants, Andolina et al. [3] explored a search agent that listens, proactively retrieves, and presents relevant information during a conversation. While the participants recognized the potential benefit, they were concerned about the lack of control over the search process done by the proactive search agent. McMillan et al. [26, 27] recorded daylong audio streams from 10 participants and observed their concerns about the unintended exposure of confidential data and possibly inaccurate projection of their personas. They also found that giving users proper control over their audio recordings and transparency over the use of inferences made based on the collected data are critical factors for the acceptance of such passive-listening technologies. They emphasized that giving users the option to examine the recordings and actions taken by the systems will encourage adoption [25]. Therefore, privacy-preserving features may become not only a competitive advantage, but an unavoidable prerequisite for the next generation of voice assistants that will always be listening.

While prior research has shed light on the challenges pertaining to always-listening systems, we are unaware of any that has directly studied how potential privacy and functionality issues impact end users' decisions to use these systems. To fill this gap, in this study, we investigate how comfortable users are to use an always-listening device, the nature of services users expect from the device, the circumstances in which those services may be considered beneficial or raise privacy concerns, and what trade-off users recognize between sharing their conversation and receiving assistance from the device.

3 METHODS

Our study aims to understand users' expectations about the services an always-listening device could offer and to examine factors that influence users' decisions to use such a device. To this end, we conducted an online survey with 178 people using the crowdsourcing platform Prolific[1] in January 2019. For the survey, we recruited participants who were English speakers, aged 18 and older, and residents of the US, who had at least a 95% approval rate on Prolific. Participants took on average 10.5 minutes to complete the survey and received $1 as compensation. Our survey and all materials were approved by relevant IRBs.

Since always-listening voice assistant devices are not yet available on the market, we used hypothetical scenarios to begin our inquiry of expectations about them. In the survey, we first presented participants with a description of an imaginary always-listening voice assistant that can provide services and recommendations based on users' conversations, without being explicitly invoked. Then we asked participants to listen to an audio recording of an approximately two-minute-long real-life conversation between two or more people in a home. (The transcript of the recording was also provided for convenience.) We asked participants to imagine that they are one of the speakers in that conversation recorded by an always-listening device. The recordings varied by conversation topic, including Health (32%), Relationships (17%), Food and Entertainment (6%), Travel (8%), Shopping (12%), Career (9%), and Others (16%). The recordings also varied in terms of the relationship between those speaking, including Significant Others (13%), Family members (25%), and Friends (62%).

After participants listened to the recording, we asked them to suggest at least three services an always-listening voice assistant may offer to them based on the conversation. Participants reported 543 different services; 517 of

[1] https://prolific.ac/

those were used in our analysis.[2] For each proposed service, we asked participants to rate, on a 5-point Likert scale, how comfortable they are with the voice assistant using the recording to provide the given service and how useful the service appears to them. We also asked participants whether, to receive this service, they would allow the voice assistant to use the full conversation, only part of it, or deny the use of the conversation entirely. Additionally, participants were asked to explain the reasons behind their decision in a free text response. We further asked them how sensitive they think the conversation in the recording was in general.

We also asked whether participants owned smart voice assistants, so that we could differentiate participant expectations' based on their experiences with similar devices. Finally, we collected demographic information on gender, age, and education level. Additionally, in order to increase data quality, we asked a few questions to check participants' attention and comprehension of the study instructions and screened out the subjects who answered incorrectly. (For details, see the survey in Appendix A.)

The sample conversations used in our study were drawn from corpora of natural conversations audio-recorded for linguistic studies. We used four public audio corpora: Santa Barbara [11], Call Friend South [6], North American English Corpus [5], and Call Home English Corpus [7]. We listened to the recordings of the conversations and extracted snippets of audio that lasted about two minutes and contained a conversation focused on a specific topic. We chose to limit the length of the conversation so that it would not impose too much cognitive load on the participants when they were recalling the conversation to answer our survey questions. Once the snippets were extracted, three researchers discussed and classified them based on the primary topic of conversation.

In the survey, we randomly assigned one of 33 audio snippets to each participant,[3] so that each conversation snippet was assigned to at least 5 participants. We used the recordings as an instrument to understand how different factors—such as the conversation topic and its sensitivity, or the relationship between speakers—affect participants' preferences. Thus, the number of participants per snippet was not chosen with the goal of achieving statistical significance *for each snippet*—instead we aimed for broad coverage of each topic and type of relationship. Therefore, we ensured both (1) that a substantial number of participants were assigned to each topic and relationship (e.g., conversations about health were assigned to 56 people, relationships to 30 people, conversations between family members to 44 people, etc.) and (2) the diversity of those conversations and topics.

There were multiple reasons for using linguistic corpora instead of participants' own conversations. First, the corpora provide conversations occurring in realistic and natural settings, while preserving participants' privacy; we did not record their own private conversations. Second, the corpora offered a variety of contexts and factors that may influence participants' responses, such as the topic of the conversation or relationship between the speakers. If we used the participants' own conversations (with their consent), it may have limited the variation in contexts and factors, or introduced selection bias. For instance, it could eliminate certain types of conversation topics (e.g., health, relationship, etc.) as the participants may not have shared that with us due to perceived sensitivity. Third, the corpora provide us the flexibility to present the same conversation to multiple participants for between-subject analysis in a controlled setting, thus increasing the internal validity of our results.

Data Analysis. We used a mixed-methods approach to analyze the results. For the quantitative analysis, we used mixed model linear and ordered logistic regressions with a random intercept per participant. First, we explored the full feature set (both dependent and independent variables) for the regression model. Then, we selected the best (reduced) models by applying a backwards elimination approach. At each step of model selection, we dropped the factor with the highest p-value greater than 0.05 and compared the models based on the Bayesian Information Criterion (BIC) value. We continued eliminating factors until the BIC stopped decreasing.

[2]We excluded 26 responses in which participants mentioned multiple services in one text field, as it was not clear which of the proposed services their responses in the follow-up survey questions were related to.

[3]We used a between-subject design to avoid cognitive overload and fatigue, ordering effects, anchoring effects, common method bias, and ensure the independence of observations, therefore increasing data quality and the validity of results.

For the qualitative analysis, one researcher performed preliminary open coding of the free text responses and developed initial codebooks classifying (1) the potential services that could be rendered by smart voice assistants per participants' suggestions, and (2) the reasons for allowing or not allowing the assistant to use full or partial recordings of conversations to provide services. The rest of the research team then discussed and finalized the initial codebook. Two independent coders used the codebooks to independently assign codes to the open-ended survey responses. The inter-rater agreement rate, as measured by the method from Kupper and Hafner [17], was on average 0.79 (min=0.71, max=0.87). Finally, the two coders discussed and resolved the coding disagreements.

Participants. Of the study participants, 53% identified themselves as female and 45% as male.[4] Participants were on average 32 years old ($SD = 11.62, min = 18, max = 66$). Most had a Bachelor's degree (34.3%) or some college education (37.5%). Others reported having completed high school (12.5%), a Master's degree (11.2%) or a Doctoral degree (2.8%); two participants did not complete high school. At the time of the survey, 42.5% of the participants owned a smart home voice assistant, such as an Amazon Echo or Google Home.

4 RESULTS

In this section, we report the results of our quantitative and qualitative analysis. We first looked at respondents' expectations for the kinds of services an always-listening voice assistant could provide based on the audio it detects. Then we explored the factors affecting participants' preferences for sharing conversation recordings in exchange for particular services from the device. We begin by analyzing the factors that we hypothesized may have an effect on participants' decisions (including service usefulness, participants' comfort in sharing a conversation, conversation sensitivity, etc.). We complement this analysis by examining the factors that emerged from the free-form responses through which participants explained the reasoning behind their decisions.

4.1 Services Expected from Always-listening Voice Assistants

To understand participants' expectations for always-listening voice assistants, we explored the services they anticipate such an assistant may provide. After presenting each participant with a randomly selected audio snippet, we asked them to propose three services that an always-listening device might offer. Participants came up with a variety of services based on the provided conversations (Table 1). These can be broadly divided into three categories: the voice assistant (1) providing information or recommendations, (2) storing and retrieving information for the user(s), and (3) carrying out actions on behalf of the user beyond just providing information.

Most frequently, participants expected to receive product or service recommendations from the device, followed by providing general information concerning a topic of the conversation (e.g., a weather forecast). A substantial portion of participants expected the device to store information from previously recorded conversations and use that to fulfill users' future requests, such as providing event reminders. As for carrying out actions, a few participants expected the device to make purchases (e.g., book a flight or play an audio book), provide driving directions when justified by the conversation's context (e.g., to show the available driving routes and directions to a place recently mentioned), and send emails or messages on behalf of the users (e.g., send a message to the conversation partner, confirming the discussed plans, or call a person mentioned in the conversation).

We observe that current voice assistants can already provide most of the services imagined by our participants (e.g., providing general or specific information). However, the always-listening voice assistant would be able to offer those services proactively, without an explicit request, and in a timely manner. Some participants emphasized an added benefit of, for example, immediately initiating communication with those mentioned in the conversation to "remove the hesitancy most people experience and can help him communicate with his brother" (P73) or to "add

[4] Three participants self-identified as neither, and one did not disclose their gender.

Table 1. Types of services an always-listening voice assistant could provide, based on participants' coded responses. Numbers in columns indicate how many times a service was mentioned in total (as each participant proposed multiple services) and how many participants mentioned the service at least once.

Type of services	Description and Examples	Services	Participants
Recommend products, places, or services	Recommend specific products, places, or services based on users' needs or interests inferred from the conversation - e.g., suggest restaurants to eat, recommend gifts for an upcoming birthday - *"A recommendation for a place to eat. This would help [those talking] find something that is casual, reasonably priced, and near their location"* (P3)	212 (41%)	113 (63.5%)
Provide general information	Provide general information or advice (without recommending a specific product or service) relevant to the topic of the conversation - e.g., advice on diet, information about symptoms of a health condition - the voice assistant could *"provide information on getting into real estate, if [the person speaking] is unhappy getting in car sales he can look into how to get into another trade"* (P108)	130 (25.1%)	85 (47.8%)
Track, store and retrieve information	Store information from the conversation and retrieve information on users' needs - e.g., add calendar event, set reminder, track when something was sent and received - *"The man would not have had to ask his wife how much their child weighed. This information would probably have been stored [in the voice assistant] as they probably talk about their child quite a bit"* (P164)	79 (15.3%)	63 (35.4%)
Create appointment/ reservation	Make reservations or appointments - e.g., schedule a meeting or appointment with a doctor - *"When a participant in a conversation mentions he or she may have an illness, the device would respond with a request to make a doctor's appointment at an appropriate facility"* (P126)	17 (3.3%)	16 (9%)
Initiate communication	Text, call or email someone on behalf of the user - e.g., contact emergency services in case an emergency situation was inferred, text a friend about running late to the movie theater - *"The Voice Assistant could offer to have the person dictate an email to send to their doctor to ask about the side effects (of a medicine), or perhaps the Voice Assistant can dial the phone right now"* (P148)	17 (3.3%)	14 (7.9%)
Provide navigation	Give users direction on how to reach a certain destination - e.g., directions on how to walk or drive to a specific park - *"The voice assistant can give the man optimal routes to his intended destination of the day"* (P90)	17 (3.3%)	15 (8.43%)
Purchase products	Purchase products or services on behalf of the user - e.g., ordering a gift for a friend's upcoming birthday - *"The Voice Assistant could immediately offer to order a pill container. Although the woman says she resents the medicine, it sounds like she realizes she needs to take it, so a pill organizer would be helpful"* (P36)	14 (2.7%)	12 (6.74%)
Provide weather information	Inform user about weather conditions - e.g., assistant could tell the user what the current weather is in his current location - *"Give more detailed upcoming weather to [speaker], who was clueless about the cold front. Offering upcoming forecasts and descriptive information about the following day's weather"* (P89)	14 (2.7%)	13 (7.3%)
Other services	There are other services participants reported that do not fall into any of the above categories. - e.g., automatically set the temperature of the house, post social media updates, or act as a therapist - *"Depressed people might be able to get help from the Voice Assistant because they would have someone to talk to. Not being alone would help even if it's just talking with the Voice Assistant"* (P99)	17 (3.3%)	17 (9.6%)

additional relevant people to the conversation" (P73), in a convenient way. Calling a doctor or emergency services in response to keywords such as "gun" or "kill oneself" was another example mentioned by the participants.

4.2 Intent to Share the Recorded Conversations

Our study participants suggested a variety of services always-listening devices might provide—but would they themselves be willing to use them? And how do they arrive at their decisions? To explore respondents' sharing preferences, we asked them whether they would permit the voice assistant to use the conversations they heard in exchange for the services they proposed. For roughly half of the services (47.16%), participants said they would deny such functionality, and for 25.04% of the services, they would allow only parts of the conversation to be used. For the rest (27.8%), participants were willing to share the entire conversation they heard to receive the service.

The preferences were not uniform: neither among participants, nor among services. A significant portion of participants (29.77%) did not want to allow the voice assistant to use audio for *any* of their suggested services. In

Table 2. Regression models for the intent to allow sharing the conversations to receive services.

Factors	Full Model				Reduced Model			
	Estimate	Std Err	t value	p value	Estimate	Std Err	t value	p value
(intercept) Fully Allow|Deny	-5.36	1.29	-3.91	.000***	-5.44	0.75	-7.25	.000***
(intercept) Part Allow|Deny	-2.89	1.27	-2.27	.024*	-3.41	0.72	-4.71	.000***
Comfort	0.7	0.11	6.55	.000***	0.68	0.1	6.95	.000***
Service Usefulness	0.65	0.15	4.46	.000***	0.61	0.14	4.48	.000***
Recording Sensitivity	-0.31	0.143	-2.165	.031*	-0.23	0.11	-1.97	.047*
Conversation Topic (Baseline=Health)								
Relationship	0.7	0.5	1.4	.162				
Travel	-0.59	0.7	-0.85	.397				
Shopping	-0.02	0.62	- 0.03	.973				
Career	0.42	0.632	0.66	.509				
Food & Entertainment	-0.35	0.74	-0.47	.637				
Service Type (Baseline=Provide general info)								
Recommend product(s)/place(s)	-0.11	0.32	-0.35	.731				
Create appointment/reservation	1.06	0.68	1.57	.116				
Initiate communication	-0.8	0.73	-1.1	.272				
Provide navigation	-0.11	0.71	-0.15	.879				
Purchase products	-0.14	0.73	-0.19	.85				
Provide weather information	-0.34	0.77	-0.44	.658				
Track, store & retrieve information	0.07	0.38	0.18	.855				
Speakers' Relationship (Baseline= Significant Other)								
Family member(s)	0.27	0.53	0.51	.614				
Friend(s)	0.32	0.63	0.52	.605				
Smart Speaker Owner	0.73	0.34	2.12	.035*	0.76	0.3	2.55	.011*
Female	-0.45	0.34	-1.34	.182				
Age	-0.003	0.02	-0.21	.835				
Education	-0.10	0.18	-0.57	.567				
Observations	517				517			
Log-Likelihood	-2123.414				-2071.794			
Akaike Information Criteria	4248.837				4145.595			
Bayesian Information Criteria	4253.025				4149.82			

*p<.05 **p<.01 ***p<0.001

contrast, 38.76% of participants allowed the use of audio for all services. Another 31.47% of participants varied their decisions based on the conversations they heard.

To gain better insight about factors that influence participants' willingness to share, we estimated the parameters in a regression model, with the decision to allow access as the dependent variable. As independent variables, we included the level of usefulness and comfort with the service, perceived sensitivity of the recording, conversation topic, service type, and demographics.

The regression analysis (Table 2) shows that perceived sensitivity of the conversation, smart speaker ownership, service usefulness, and level of comfort have a significant effect on participants' intentions to allow or deny access to the conversation. However, conversation topics, service types, speakers' relationship, and participants' demographics appear to have no significant effect on their sharing intent among our sample.

Participants were more willing to allow the voice assistant to use the conversations if they found the services useful and were generally comfortable with this functionality. Unsurprisingly, as the sensitivity of the conversation increases, participants become more inclined to deny access to the recording. We also found that perceived comfort with the service and usefulness of the service are better predictors of participants' sharing intent than the perceived sensitivity of the conversation.

Familiarity with existing voice assistants also played a role: current smart speaker owners were more willing to allow the use of conversations in exchange for services. These results suggest that potential users consider the trade-offs between service usefulness, their personal level of comfort, and the conversation sensitivity when deciding whether or not to share their recordings.

Table 3. Regression models for the perceived usefulness of the services (Usefulness).

Factors	Full Model				Reduced Model			
	Estimate	Std Err	t value	p value	Estimate	Std Err	t value	p value
(intercept)	3.31	0.41	8.05	.000***	3.7	0.27	13.64	.000***
Comfort	0.20	0.04	5.61	.000***	0.21	0.03	6.14	.000***
Recording Sensitivity	0.09	0.05	1.69	.092				
Conversation Topic (Baseline=Health)								
Relationship	-0.11	0.18	-0.59	.557				
Travel	0.11	0.25	0.46	.647				
Shopping	0.19	0.23	0.82	.412				
Career	-0.03	0.23	-0.12	.906				
Food & Entertainment	0.17	0.27	0.65	.518				
Service Type (Baseline=Provide general info)								
Recommend product(s)/places(s)	0.04	0.11	0.363	.717	0.06	0.10	0.575	.566
Create appointment/reservation	0.31	0.23	1.36	.174	0.36	0.22	1.62	.105
Initiate communication	0.53	0.23	2.32	.021*	0.54	0.23	2.38	.018*
Provide navigation	0.1	0.24	0.42	.674	0.12	0.23	0.54	.59
Purchase products	0.58	0.26	2.25	.025*	0.61	0.25	2.42	.016*
Provide weather information	0.30	0.25	1.22	.224	0.26	0.24	1.05	.294
Track, store & retrieve information	0.4	0.13	3.1	.002**	0.38	0.13	3.06	.002**
Speakers' Relationship (Baseline= Significant Other)								
Family member(s)	-0.2	0.22	-0.89	.38				
Friend(s)	-0.03	0.2	-0.15	.883				
Smart Speaker Owner	0.08	0.12	0.63	.533				
Female	-0.03	0.12	-0.26	.792	-0.06	0.12	-0.48	.630
Education	-0.14	0.06	-2.24	.025*	-0.13	0.06	-2.15	.032*
Age	0.004	0.01	0.69	.493				
Observations	517				517			
Log-Likelihood	-690.904				-680.203			
Akaike Information Criteria	1385.833				1364.430			
Bayesian Information Criteria	1394.209				1372.851			

*p<.05 **p<.01 ***p<0.001

4.3 Perceived Usefulness of the Services

Our initial analysis concluded that the benefits associated with the services play a major role in participants' decisions to share conversation in exchange for such a service—but what makes the service beneficial? Under which circumstances do people find it useful to receive a service from the voice assistant? We examined participants' survey responses about the perceived usefulness of the services and the factors potentially affecting it.

Most of the services were deemed to be slightly (46.4%) or very (30.9%) useful. This is not altogether surprising, because participants rated the services they suggested themselves. Mixed effects linear regression analysis (see Table 3) revealed that the perceived usefulness of a service is significantly influenced by participants' comfort with the service, particular service types (such as communication initiation; product purchasing; tracking, storage and retrieval of information), and participants' education level. In contrast, conversation topic and sensitivity, speaker relationship, ownership of smart speakers, age, and gender did not have a significant effect.

As stated earlier, the type of service also has a significant impact on the perceived usefulness of the service as well. While 'providing general information' is a popular service, it is not considered as the most useful. Participants find the device initiating communication, making a purchase, or storing and retrieving information more beneficial than simply providing information or recommendations. It is possible that participants may find searching for information by themselves easy, while actions like purchasing products require more effort, and assistance is consequently welcomed. Moreover, participants may compare a voice assistant's recommendations to hearing advertisements, which are often unwanted, especially from smart speakers [22].

Table 4. Regression models for the perceived level of comfort (Comfort).

Factors	Full Model				Reduced Model			
	Estimate	Std Err	t value	p value	Estimate	Std Err	t value	p value
(intercept)	2.84	0.63	4.53	.000***	2.53	0.44	5.8	.000***
Service Usefulness	0.24	0.05	4.67	.000***	0.24	0.05	4.8	.000***
Recording Sensitivity	-0.1	0.08	-1.14	.254				
Conversation Topic (Baseline=Health)								
Relationship	-0.79	0.28	-2.88	.004**	-0.88	0.27	-3.31	.001**
Travel	-0.41	0.37	-1.01	.277	-0.42	0.35	-1.2	.231
Shopping	-0.2	0.35	-0.57	.572	-0.09	0.31	-0.29	.772
Career	0.39	0.35	1.1	.272	0.33	0.34	0.87	.33
Food & Entertainment	-0.23	0.41	-0.56	.578	-0.17	0.39	-1.45	.655
Service Type (Baseline=Provide general info)								
Recommend product(s)/places(s)	-0.14	0.12	-1.21	.225	-0.14	0.12	-1.21	.226
Create appointment/reservation	-.59	0.24	-2.42	.016*	-.59	0.24	-2.42	.016*
Initiate communication	-0.04	0.25	-0.17	863	-0.04	0.25	-0.17	.868
Provide navigation	0.20	0.26	0.79	.432	0.19	0.26	0.75	.455
Purchase products	-0.52	0.28	-1.84	.066	-0.52	0.28	-1.83	.068
Provide weather information	0.37	0.27	1.35	.178	0.38	0.27	1.39	.166
Track, store & retrieve information	0.02	0.14	0.12	.906	0.02	0.14	0.11	.91
Speakers' Relationship (Baseline= Significant Other)								
Family member(s)	-0.22	0.34	-0.65	.517				
Friend(s)	-0.12	0.29	-0.37	710				
Smart Speaker Owner	0.53	0.19	2.81	.005**	0.56	0.19	3.05	.002**
Female	-0.19	0.19	-1.01	.313				
Education	-0.21	0.097	-2.19	.029*	-0.19	0.09	-2.06	.04*
Age	0.014	0.01	1.51	.133				
Observations	517				517			
Log-Likelihood	-772.417				-767.929			
Akaike Information Criteria	1548.859				1539.882			
Bayesian Information Criteria	1557.235				1548.283			

*p<.05 **p<.01 ***p<0.001

4.4 Perceived Level of Comfort

To further advance the analysis, we explored the following questions: how comfortable are participants with the assistant using the recorded conversation to provide services? What factors make participants more or less comfortable sharing the recording with the device?

Slightly more than half of the time (52.8%), participants were uncomfortable sharing the conversation to receive a service. Participants were most comfortable (58.8% of the time the service was mentioned) sharing the recording in exchange for the device initiating communication on their behalf. Participants might think such a service would help users quickly get in touch with someone, especially in an emergency, which made them more comfortable with sharing the conversation. Participants were most uncomfortable sharing the recording for receiving suggestions and recommendations from the voice assistant (62.3% of the time the service was mentioned). Participants may not find this type of offering useful enough to share their conversations.

The regression analysis (see Table 4) showed that participants' comfort with receiving a service depends on perceived usefulness and type of service, conversation topic, smart speaker ownership, and participants' education levels. However, sensitivity of the conversation, speakers' relationships, gender, and age appear to have no effect among our sample.

Specifically, participants' comfort with sharing the conversation increases with the perceived usefulness of the service. Participants were significantly less comfortable with the voice assistant creating an appointment or reservation on behalf of the user, as compared to the assistant providing general information. Among different conversational topics, participants were significantly less comfortable sharing a conversation about relationships with the device than conversations about health. However, the sensitivity of the recorded conversation did not significantly relate to the level of comfort. Rather, participants seemed to make trade-offs between the benefit

Table 5. Regression models for the perceived sensitivity of the conversation recordings (Sensitivity).

Factors	Full Model				Reduced Model			
	Estimate	Std Err	t value	p value	Estimate	Std Err	t value	p value
(intercept)	4.46	0.46	9.54	.000***	4.68	0.21	22.71	.000***
Conversation Topic (Baseline=Health)								
Relationship	0.36	0.27	1.32	.188	0.3	0.27	1.11	.268
Travel	-1.00	0.36	-2.80	.006**	-1.12	0.34	-3.26	.001**
Shopping	-1.54	0.32	-4.86	.000***	-1.5	0.3	-4.96	.000***
Career	-0.37	0.35	-1.07	.287	-0.47	0.34	-1.39	.166
Food & Entertainment	-1.12	0.395	-2.82	.005**	-1.21	0.39	-3.15	.002**
Speakers' Relationship (Baseline= Significant Other)								
Family member(s)	-0.29	0.34	-0.85	.399				
Friend(s)	-0.34	0.29	-1.16	.247				
Smart Speaker Owner	-0.28	0.19	-1.49	.139	-0.29	0.19	-1.57	.119
Female	-0.29	0.18	-1.61	.109	-0.3	0.18	-1.63	.106
Age	0.01	0.01	0.87	.386				
Education	0.07	0.1	0.71	.476				
Observations	178				178			
Log-Likelihood	-284.031				-279.141			
Akaike Information Criteria	570.086				560.306			
Bayesian Information Criteria	573.162				563.406			

*p<.05 **p<.01 ***p<0.001

received from the services and the sensitivity of the conversation—with benefits often outweighing the sensitivity. In 27.5% of the cases, participants reported being comfortable sharing the recordings in exchange for a service, regardless of whether they perceived the conversation in the conversation as somewhat or very sensitive.

We also found that less educated participants and current owners of smart speakers were more comfortable allowing the use of conversations by the always-listening devices. This suggests that current users of smart speakers may be already accustomed to receiving many of the suggested services, which made them more open to the idea of automating this process.

4.5 Perceived Sensitivity of the Conversations

Earlier, we found that the sensitivity of the conversation is a significant predictor of participants' decisions to allow the voice assistant access to their conversation. We wanted to further understand: what makes a conversation less or more sensitive to participants?

Overall, participants found the conversations they heard very (43.8%) or somewhat (30.3%) sensitive. Regression analysis (Table 5) demonstrates that the topic of the conversation is the only factor in our model that has a significant impact on the perceived sensitivity of the recording. Specifically, conversations about travel, shopping, and food and entertainment were considered significantly less sensitive than conversations about health. However, conversations about relationships and careers were not significantly different in sensitivity than discussions about health. Relationships between the speakers, ownership of a voice assistant device, and demographic factors did not significantly affect the perceived sensitivity of a given conversation.

Our findings demonstrate that the sensitivity of a particular recording partially depends on the topic of the conversation, which in turn influences participants' sharing intents. From a device manufacturer perspective, knowledge about what topics are considered sensitive can be used for topic-based filtering [20] to avoid disclosure of embarrassing or private information in an always-listening environment.

4.6 Reasons Explaining the Intent to Share the Recorded Conversations

We predicted that a set of factors (e.g., service usefulness, comfort level, conversation sensitivity, etc.) could influence participants' intentions to allow the assistant access to the conversations. However, as this is an exploratory study, we also wanted to investigate other factors that may affect participants' preferences. Thus, we asked participants to provide a free text response about the rationale for their intentions to allow or deny sharing the conversation for each of the services they proposed.

We classified participants' responses into several categories according to the procedure described in §3. We identified seven main factors affecting participants' decisions to allow or decline the use of their conversations. The factors (in decreasing order of frequency) were: privacy concerns, perceived benefits of the services, expected performance of the voice assistant, agency, interruptions, security concerns, and general attitude towards voice assistants.[5] While the impact of benefits and privacy concerns is in line with our initial hypotheses about usefulness, comfort, and sensitivity, we identified several other reasons that are worth exploring in future work.

Privacy. Almost 3 in 4 participants (n = 129, 72.5%) mentioned privacy-related aspects playing a role in their preferences. This led 38 participants (21.4%) to decline the services. For example, P104 said: *"[It is] encroaching on privacy to an insane degree. Some things are meant to stay private. Always-listening? Are you kidding me?"* In addition to the intrusiveness of the device in general, participants were concerned with the invasive or private nature of specific services (n = 19, 10.7%) and sensitivity of the recorded conversations (n = 21, 11.8%).

We gave participants the options to share none or only part of a conversation with the device. Participants concerned about the sensitivity of the conversation were more willing to avoid parts that may disclose sensitive information (n = 14, 7.8%), than to completely block access to the conversation (n = 7, 3.9%). For example, P60 mentioned: *"Something about the voice assistant noting my travel habits or the locations of family members is just too specific and intrusive for myself to be personally comfortable with."* On the other hand, when participants perceived the service (n = 13, 7.3%) or the conversation (n = 11, 6.2%) as ordinary and non-intrusive, they were willing to allow access to the full conversation. Participants considered a service private when the voice assistant required access to some personal information to provide that service. For example, P67 said: *"I do not want a voice assistant to schedule things for me, it implicitly means that some other entity has access to my daily plans."*

In addition, three participants said they would deny use of the conversations for any service because they do not trust the manufacturer or service provider to respect their privacy. In contrast, trust towards these entities led four participants to allow the use of the full conversations for any service. Participants who allowed partial use of the conversations or allowed only some of the services did not mention trust as a reason for their decisions. This suggests that trust is one of the fundamental and non-negotiable prerequisites for the adoption of always-listening smart assistants. If people do not trust the device manufacturer or service provider, they are likely to completely refuse to use it. This is exemplified by poor sales figures for the Facebook Portal device: "after the company's many privacy sins, people are apparently hesitant to put a Facebook device with a camera in their living rooms" [41].

Benefits. The perceived benefits of the services provided by the always-listening voice assistant are another driving factor for the intentions to allow service provision. For instance, P107 appreciated the voice assistant recommending her recipes: *"Even if I don't make any of the recipes, perhaps hearing about low-salt recipes would make me more open to altering my diet."* Participants considered a service useful when it saves time or effort, informs about topics they are interested in, and, overall, increases their quality of life. For instance, P113 said: *"[The device] could tell her schedule or add the different things they talked about, so they don't forget anything. Would be very helpful because often keeping track of a schedule is difficult and things get forgotten or mis-remembered."*

Participants (n = 51, 28.7%) were willing to share entire conversations if they perceived a service as useful or beneficial. In contrast, participants said they would limit the device's access to their conversation and share only part of it (n = 20, 11.2%) or deny the service entirely (n = 14, 7.9%) if they were not interested in the service. Thus,

[5]While reasons related to benefits and privacy attitudes could have been in part driven by the survey questions, free-text responses revealed additional insights about benefits and privacy concerns. For instance, future research should distinguish clearly between the privacy issues associated with device operations (i.e., how it collects and processes the data), the conversations to be shared, and the services to be offered.

while the currently available voice assistants provide services upon explicit request, proactively provisioning services by devices will require additional effort in making sure those services are desired and timely.

Performance and Relevance. Some participants (n = 49, 27.5%) doubted the device's ability to accurately infer users' needs and effectively deliver relevant services. This led 23 participants (12.9%) to express an intention to deny service provision. Fifteen (8.4%) of them specifically said that either they would perform those tasks better by themselves or that human oversight is absolutely necessary to effectively carry out those tasks. For example, P81 said: *"Scheduling a dentist appointment is something I feel would need a full human involvement to look for the dental office you would like and describe your problem to them."*

Participants also had different opinions on the amount of information required for the voice assistant to complete the service efficiently. Twelve of them (6.7%) decided to share the entire conversation, as they believed it would improve the accuracy of the device's performance, as well as the relevance and quality of the provided services. For instance, P164 said: *"If I'm already allowing the service, I'd want it to be as accurate as possible. The more information, the better."* In contrast, some others (n = 14, 7.7%) believed the device should automatically recognize and accurately infer what information is relevant for a specific service and then should only have access to this information, discarding the irrelevant parts of the recording.

Agency. Fifteen participants (8.4%) mentioned their discomfort with a voice assistant proactively making decisions and actions on their behalf. In contrast to the participants who believed they would perform the services better themselves, respondents in this category were not worried about the quality as much as they did about the loss of agency and control over the data collection and service delivery. This led seven participants (3.9%) to deny the service completely. In contrast, eight others (4.5%) allowed the service, but stated they would need the assistant to explicitly ask for permission before carrying it out, especially if the service was related to contacting someone or purchasing a product. For example, P170 stated: *"I would not want flowers to be ordered without me seeing them and researching the company first. I wouldn't mind seeing flower shops ads."* For these participants having controls over the service provision was not an amenity, but an essential prerequisite for accepting the service.

Interruptions. Five participants (2.9%) disliked the idea of potential interruptions to a conversation caused by unprompted suggestions from the voice assistant. One person denied access for all proposed services based on this concern. He (P142) mentioned: *"I don't need constant suggestions for every little thing I talk about. I may not be ready to actually do anything about it."*

Security. Only two participants (1.1%) explicitly mentioned concerns about potential security vulnerabilities of the always-listening assistant as the reason for denying services in the presented hypothetical situations. Specifically, they were concerned about potential "hacking" of the assistant—either as a security breach happening to a third party or as a result of a company employee's or contractor's abuse of their access privileges—and the resulting unauthorized access to their private conversations.

General attitude. Generally unfavorable attitudes towards existing, keyword-triggered, voice assistants motivated the intentions of three participants (1.7%) to deny access for all services. As these participants did not elaborate on why they disliked voice assistance technologies, future work may elucidate their reasons.

In summary, we found that privacy concerns associated with always-listening functionalities, doubts about vendors' intentions to protect users' privacy, interruptions, security risks, and general dislike of voice assistants need to be addressed as rather fundamental, non-negotiable prerequisites for adoption of always-listening devices. On the other hand, privacy concerns associated with the sensitivity of specific conversation topics and services, loss of agency, the lack of perceived usefulness or relevance, and concerns about performance issues are context-specific. They therefore require granular controls, a topic researchers and designers should explore in more detail.

It is possible that specific concerns implicitly affect the general concerns; therefore, addressing known issues may reduce and mitigate the fundamental adoption barriers.

5 DISCUSSION

The core objective of our work is to understand user expectations for future always-listening voice assistants and explore the factors affecting them. Using quantitative analysis, we showed that factors like sensitivity, comfort, and usefulness can drive users' intentions to allow or decline the use of such devices. To advance our understanding, we used qualitative analysis to explore participants' concerns about always-listening devices, including privacy, trust, agency, and security. Here we discuss the context of these observations and what these observations mean for the future of always-listening devices and research.

5.1 Use Cases for Always-Listening Devices and the Challenges of Providing Services

In light of the rapid development of voice assistant technologies [13, 39], we anticipate that always-listening devices will arise in the near future. We therefore set out to understand the services people expect these devices to provide. Our results can guide system designers in understanding what services the market might demand and help researchers identify areas where current technology may not adequately meet consumers' needs.

Our respondents suggested a wide range of services, many of which—at a high level—are already available from keyword-triggered voice assistants (e.g., providing information or making purchases). However, in contrast to keyword-activated devices, seamlessly and autonomously carrying out these actions, in the way expected from a truly intelligent always-listening assistant, as suggested by participants, will be technically challenging.

One such challenge is determining the recipient of a service. In keyword-triggered assistants, it can generally be assumed that the person addressing the device is requesting the service for themselves. But an assistant trying to passively infer actions from conversations will need to handle semantic ambiguity. For example, if someone says *"oh yeah, it would be bad to forget that,"* the assistant could offer a reminder, but it may not be immediately clear who should receive that reminder—the speaker herself or one of the interlocutors.

In addition to being targeted to the right person, services must also be contextually relevant. Because in the keyword-triggered assistants, the speaker pro-actively requests the service, its relevance for the speaker is ensured and not dependent on the context. The always-listening devices will have to first infer the context, and then identify relevant services; both processes are error prone. For example, in our study, some participants rejected services that were ostensibly relevant (e.g., a recommendation of baby products following a conversation about someone being pregnant) because they did not match the broader focus of the conversation (preparing dinner, in this example). To meet these expectations, a sophisticated understanding of the conversation and its context is required, which remains a challenge for current NLP systems [28].

Furthering this challenge is the context-dependent nature of users' expectations regarding consent and control. For instance, while participants in our study preferred to be explicitly asked for consent in most situations (like confirming a purchase or contacting a friend), in an emergency situation they may be willing to bypass that in exchange for greater safety. Obtaining explicit consent for virtually every action would be conflicting with the core feature of always-listening devices—providing seamless automated assistance. Although in emergency situations, keyword-activated devices face similar challenges of not being able to contact emergency services [14], in other situations, obtaining consent is less problematic for the keyword-activated devices, as consent is often implied by the very fact of initiating the request via a wake-word.

Advancement in speech recognition and semantic analysis may be able to improve the accuracy of content recognition to address the challenges outlined above. While fully automatic systems are under development, a semi-automatic approach may be adopted, where to minimize the type I and type II errors, technology developers would incorporate mechanisms that would confirm the action and recipient before delivering a service.

Potential interruption of the conversation caused by the always-listening voice assistant was another concern, unique to this class of devices, mentioned by a few participants. To minimize interruption, the voice assistant could leverage a second output channel, such as a mobile phone or a smart display, instead of interrupting an ongoing conversation by speaking to the user. The device may begin by presenting proposed actions or suggestions on the screen, especially in the initial training phase. Once the device is trained to meet a satisfactory level of accuracy, a more immediate channel, such as audio, may be used to offer a service.

5.2 The Trade-Offs between Privacy and Utility

Prior studies of Internet of Things scenarios have found substantial privacy concerns among participants in hypothetical situations that concerned the collection and usage of voice data [19, 30]. Yet, these devices—voice assistants—are now proliferating. One likely explanation might be that users grew more comfortable with devices as they experienced real benefits in day-to-day life [33]. Our data support this hypothesis. We found that current owners of voice assistants were more likely to allow an always-listening device to use a recorded conversation than non-owners, thus making them more likely to adopt even more privacy-invasive future devices.

Although the perceived benefits might outweigh privacy concerns for some users, this does not mean that with increasing familiarity the concerns about this new technology will disappear. Past research has shown that people hesitate to adopt new technologies if they perceive them as too invasive; they may also find other *ad hoc* coping strategies, such as unplugging the device or limiting their usage, to compensate for the absence of appropriate privacy controls [1, 18].

The tension between privacy and benefits will be more pronounced with always-listening voice assistants, as private information will be collected with substantially less user control and consent, compared to keyword-activated interactions. For instance, participants found a booking/reservation service useful, but expressed discomfort about sharing the conversation in exchange for this type of service roughly half (47%) of the time. Furthermore, some of the respondents noted that the assistant's mistakes could have a tangible effect on them, such as financial damage if the voice assistant ordered the wrong product.

This calls for designing necessary privacy frameworks and controls to release the users from the potential cognitive tension and to anticipate unintended device use cases. One way to reduce this tension is to provide users with effective auditing options for their interactions with the device. Through this interface, users could review recorded conversations along with the suggested services, and decide whether they would like this information to be recorded and these services to be offered to them in the future, thereby establishing or amending the rules for the assistant to follow.

Visual or auditory indicators providing users with feedback about the voice assistant's current activity status could also be useful in mitigating privacy concerns. In the keyword-activated devices, such indicators signal that the device has recognized a wake-word, is recording, and ready to receive a request. It is useless to indicate the recording in the always-listening device, as it is always recording. Instead, it could light up or make a sound when it detects a new voice or the voice of a child (since recordings of children are considered more sensitive [22]), or when it recognizes a service that is relevant to the current part of the conversation. On the other hand, these indicators might be too easily overlooked, making such solutions inadequate. Future work should therefore explore the contexts in which such indicators would be effective.

5.3 Sharing Preferences are Nuanced

An always-listening voice assistant can provide exceptional convenience to users. Participants emphasized the value they found in increasing the efficiency of information retrieval and reducing the cognitive load of tracking daily responsibilities. Always-listening devices, however, could listen to sensitive conversations when users least expect them to do so. Therefore, they are more likely to violate users' privacy expectations, as compared to current keyword-based devices. Indeed, many participants raised privacy concerns associated with the continuous listening of their conversations. Other concerns about always-listening voice assistants were common to current keyword-triggered devices as well, for instance, lack of trust in the manufacturer's ability and intent to respect users' privacy preferences [18].

The willingness to accept services from an always-listening voice assistant was not uniform across participants. Participants split into three fairly equal groups: those who expressed an intent to allow all services, deny all services, or allow some and deny other services. Prior research has also observed similar user distributions for sensitive data sharing on smartphones [43]. Participants who are willing to allow or deny access in all cases rely on rather fundamental reasons and are less likely to take contextual factors into account. However, a large portion of participants had more granular preferences, and we wanted to understand how these participants make their decisions. Based on our data, we found a few concrete factors that might help device platforms better meet this group's privacy expectations. Among participants with nuanced preferences, the sensitivity of the conversation was important. Moreover, we found that conversation topics, such as health and relationships, could significantly influence the perceived sensitivity of a conversation. These findings are in line with prior work [9]. Existing keyword-activated voice assistants do not offer mechanisms to filter out specific types of information and prevent them from being sent over the Internet to the vendors' servers.

To address these needs, devices could systematically detect the conversation topic by using techniques such as Topic Modeling [37, 40]. Designers could allow users to blacklist certain conversation topics they perceived as sensitive, or treat them with special caution (e.g., delete such recordings or ask for users' explicit consent to use them). However, the voice assistant would need a more nuanced understanding of the conversation to effectively measure its sensitivity. For instance, a conversation about shopping is significantly less sensitive than one about health. Yet, a conversation about buying a pregnancy test may be perceived as more sensitive than a conversation about catching a cold. Some participants also suggested deploying a whitelisting approach, whereby the voice assistant would identify only the *relevant* parts of the conversation, in order to provide a specific service, while discarding the irrelevant parts.

Prior research has shown that users of today's smart speakers have different privacy expectations, depending on whether they or someone else—such as a child or guest—is speaking [22]. We therefore hypothesized that factors related to the speakers and subjects in the conversations might influence privacy expectations. While our results suggest a trend in which a conversation between significant others may be perceived as more sensitive than a conversation between friends and family, this result was not statistically significant, potentially due to an insufficient number of observations. Future research is needed to further investigate this and other factors affecting users' expectations regarding the sensitivity of a particular conversation. For instance, the Contextual Integrity (CI) framework provides a baseline for taking context into account in meeting consumers' nuanced privacy expectations [31]. The CI framework considers how information flows in a particular contextual situation to determine whether or not it matches societal expectations, and may be useful in designing user studies and contextually-aware privacy controls. This groundwork, in turn, could help future platform designers to develop and implement usable control systems and consent mechanisms, e.g., how to ask users for permission to share their conversations, when to ask for permission, and how to design privacy controls centered around these factors.

5.4 Privacy in a Multi-user Setting

Unlike a typical smartphone, a smart speaker serves multiple people at the same time. Researchers have started exploring the attendant implications for keyword-activated assistants [18, 22, 44]. Our findings echo concerns expressed in these studies, but suggest that some issues will be more severe in an always-listening environment.

While keyword-activated devices can also be used by multiple users, to not be recorded, people can simply withdraw from interacting with the assistant by not using the wake-word. It will be harder to remain anonymous from always-listening devices; even to prohibit the recording of certain household members, the device would need to be able to detect and recognize their voices.

Furthermore, the always-listening voice assistant would need to manage access to the data collected about multiple household members. Participants mentioned that the voice assistant could support storing and retrieving of user information. But who should have access to this information once it has been stored? In certain cases, the data may be associated with external accounts for which users may have distinct sharing preferences. For example, appointments would be stored in a calendar, for which users can already control the visibility of events. In these cases, the assistant could adhere to the data sharing settings contained within that application. This suggestion may be implemented in both keyword-activated and always-listening voice assistants.

Moreover, there may be other data types—not used by currently offered services, but required for the novel ones in the future—that would require new storage, access, and sharing settings. For instance, one participant expected the voice assistant to *"remember the [symptoms of a] previous ectopic [pregnancy]. This way, the service can have stored that this pregnancy feeling may be another ectopic instead of an actual pregnancy [ahead of time]"* (P88). If an assistant decides to keep track of this information, the system designer would need to also develop appropriate access-control settings, since the user may not expect this information to be accessible to her children. Thus, always-listening voice assistant needs to be able to identify and properly label what information to store and whom it is about, where to store it, and who in the household is allowed access to it.

Another challenge is resolving contradictions between different users' personal policies and preferences. Two (or more) people holding a conversation may have different disclosure preferences. Whose opinion should the device prioritize when deciding whether to share a particular conversation? Should it default to the owner's preferences? Should it default to the most privacy-protective preference? While also relevant for the keyword-activated devices, this issue is more prominent in an always-listening environment, because users are actively participating in conversations all the time, whereas a keyword-activated voice assistant listens to the audio only when explicitly invoked. Therefore, additional privacy management settings are especially necessary for always-listening voice assistants so that users can set their individual preferences and choose how to handle sharing data in case the preferences conflict.

It is also challenging, for both the keyword-activated and always-listening voice assistants, to identify when a service that is appropriate in a single-user setting becomes problematic in a multi-user environment [18]. For instance, users may be comfortable with the voice assistant reminding them out loud about mowing the lawn but may be very uncomfortable being reminded, in front of others, about taking birth control pills.

It is possible that addressing this challenge is in fact easier with an always-listening voice assistant, than with a keyword-activated one, at least in certain situations. By paying attention to the background and context, the device can detect if there is another person present in the room and change the method of delivery for sensitive services; for instance, it could send a mobile notification instead of saying something out loud. Device designers could also incorporate a "training phase," in which a family would teach the assistant about its norms and preferences, such as what topics are appropriate for which family members.

5.5 Limitations

As with many online surveys, our study may not be fully representative of the US population; for example, our sample is, on average, more educated. The limited sample size is another limitation. A future large scale study could validate our findings across a wider population.

Always-listening voice assistants are not yet commercially available, but based on recent patent filings [2, 12, 13, 39], they are clearly on the horizon. We wanted to elicit participants' *expectations* about services these always-listening assistants may provide. Therefore, instead of surveying users' actual experiences or offering well-defined hypothetical scenarios via vignette studies, we relied on our participants imagining the devices' capabilities. Providing a list of services prepared by the researchers could bias participants' responses and lead to functional fixedness. We acknowledge, however, that our respondents may not fully understand the expanse of services that such a system can offer. A natural future step is to study users' expectations, concerns, preferences and the relevant factors *in situ*.

Another limitation is that respondents were presented with conversations they did not participate in themselves. Therefore, the degree of empathy and relevance felt for the speakers might have varied between participants, affecting their expectations for the assistant. We recognize that our study is hypothetical and may not capture consumers' expectations and concerns that may arise when such a device is adopted in real life and their personal conversations are being be used. However, assessing how relevant a particular personal audio snippet is to a respondent in the survey would require gathering a significant amount of personal information. Alternatively, asking participants to provide their own conversation recordings could introduce self-selection and topic-selection bias. Thus, we decided against this. Using pre-selected conversation snippets, we managed to explore a diverse set of conversation attributes and topics, in a controlled manner, without compromising participants' privacy. Future field studies could validate the results in a more ecologically-valid setting. For instance, future studies could deploy a passive listening device in participants' homes to collect a corpus of their own conversations and investigate (e.g., using the experience-sampling method) their contextually-aware expectations and concerns around those recordings.

6 CONCLUSION AND FUTURE WORK

In this work, we explored user preferences and expectations surrounding a next generation of voice assistants that could passively listen to people's conversations and proactively provide assistance. Our quantitative and qualitative analysis identified both expectations and challenges that require further investigation. As with current voice assistants, privacy issues remain a key concern that may limit users' willingness to even consider an always-listening voice assistant. Furthermore, our results highlight the need for filtering and preference controls to reduce the amount of information needed by the voice assistant to provide services. Yet, our results also indicate that users can see the distinct benefits of more personalized and contextual services and may be willing to share even sensitive conversations to receive those benefits. There is still a need for further research to identify and develop the interactive features and privacy protections and controls that would enable users to comfortably enjoy the benefits of an always-listening voice assistant.

ACKNOWLEDGMENTS

We thank our participants and anonymous reviewers. This work was supported by a grant from the Center for Long-Term Cybersecurity (CLTC) at U.C. Berkeley; National Science Foundation grants CNS-1514211, CNS-1528070, and CNS-1801501; the National Security Agency's Science of Security program; the generosity of Cisco and Mozilla; and the Wallenberg Artificial Intelligence, Autonomous Systems and Software Program (WASP), funded by the Knut and Alice Wallenberg Foundation. Opinions, findings, and conclusions are those of the authors and do not necessarily reflect the views of the funders.

REFERENCES

[1] 2019. "I don't own the data": End User Perceptions of Smart Home Device Data Practices and Risks. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. USENIX Association, Santa Clara, CA. https://www.usenix.org/conference/soups2019/presentation/tabassum

[2] Alexa Feature Help. [n. d.]. How Does Drop In Work? https://www.amazon.com/gp/help/customer/display.html?nodeId=202153130 (last accessed on August 8, 2019).

[3] Salvatore Andolina, Valeria Orso, Hendrik Schneider, Khalil Klouche, Tuukka Ruotsalo, Luciano Gamberini, and Giulio Jacucci. 2018. Investigating Proactive Search Support in Conversations. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. ACM, New York, NY, USA, 1295–1307.

[4] Barry Brown, Moira Mcgregor, and Donald McMillan. 2015. Searchable Objects. 508–517.

[5] Alexandra Canavan and George Zipperlen. 1996. CALLFRIEND American English-Non-Southern Dialect LDC96S46. Web Download. Philadelphia: Linguistic Data Consortium. https://ca.talkbank.org/access/CallFriend/eng-n.html. Accessed: 2019-03-19.

[6] Alexandra Canavan and George Zipperlen. 1996. CALLFRIEND American English-Southern Dialect LDC96S47. Web Download. Philadelphia: Linguistic Data Consortium. https://ca.talkbank.org/access/CallFriend/eng-s.html. Accessed: 2019-03-19.

[7] David Graff Canavan, Alexandra and George Zipperlen. 1997. CALLHOME American English Speech LDC97S42. Web Download. Philadelphia: Linguistic Data Consortium. https://ca.talkbank.org/access/CallHome/eng.html. Accessed: 2019-03-19.

[8] Juan Pablo Carrascal, Rodrigo De Oliveira, and Mauro Cherubini. 2015. To Call or to Recall? That's the Research Question. *ACM Trans. Comput.-Hum. Interact.* 22, 1, Article 4 (March 2015), 30 pages. https://doi.org/10.1145/2656211

[9] James Clawson, Jessica A. Pater, Andrew D. Miller, Elizabeth D. Mynatt, and Lena Mamykina. 2015. No Longer Wearing: Investigating the Abandonment of Personal Health-tracking Technologies on Craigslist. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 647–658.

[10] Benjamin R. Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. "What Can I Help You with?": Infrequent Users' Experiences of Intelligent Personal Assistants. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '17)*. ACM, New York, NY, USA, Article 43, 12 pages.

[11] Wallace L. Chafe Charles Meyer Sandra A. Thompson Robert Englebretson Du Bois, John W. and Nii Martey. 2000-2005. Santa Barbara corpus of spoken American English, Parts 1-4. Philadelphia: Linguistic Data Consortium. http://www.linguistics.ucsb.edu/research/santa-barbara-corpus. Accessed: 2019-03-19.

[12] Echo Device Help. 2019. Turn on Follow-Up Mode. https://www.amazon.com/gp/help/customer/display.html?nodeId=202201630.

[13] Google Home Help. 2019. Have a conversation with Google Home. https://support.google.com/googlehome/answer/7685981?hl=en.

[14] Jefferson Graham. 2017. 'Alexa, call 911' won't work. Here's what will. *USA Today* (July 2017). https://www.usatoday.com/story/tech/talkingtech/2017/07/19/alexa-cant-dial-911-but-google-alexa-and-siri-can-get-you-help/486075001/

[15] Jonathan Kilgour, Jean Carletta, and Steve Renals. 2010. The Ambient Spotlight: Queryless Desktop Search from Meeting Speech. In *Proceedings of the 2010 International Workshop on Searching Spontaneous Conversational Speech (SSCS '10)*. ACM, New York, NY, USA, 49–52.

[16] Bastian Könings, Florian Schaub, Michael Weber, and Frank Kargl. 2010. Towards territorial privacy in smart environments. In *2010 AAAI Spring Symposium Series*.

[17] Lawrence L. Kupper and Kerry B. Hafner. 1989. On Assessing Interrater Agreement for Multiple Attribute Responses. *Biometrics* 45, 3 (Sept. 1989), 957. https://doi.org/10.2307/2531695

[18] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. 2018. Alexa, Are You Listening?: Privacy Perceptions, Concerns and Privacy-seeking Behaviors with Smart Speakers. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 102 (Nov. 2018), 31 pages.

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

[20] Yuefeng Li, Xujuan Zhou, Peter Bruza, Yue Xu, and Raymond YK Lau. 2008. A two-stage text mining model for information filtering. In *Proceedings of the 17th ACM conference on Information and knowledge management*. ACM, 1023–1032.

[21] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf Between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 5286–5297.

[22] Nathan Malkin, Joe Deatrick, Allen Tong, Primal Wijesekera, Serge Egelman, and David Wagner. 2019. Privacy Attitudes of Smart Speaker Users. *PoPETs* 2019, 4 (2019), 250–271. https://petsymposium.org/2019/files/papers/issue4/popets-2019-0068.pdf

[23] Nathan Malkin, Serge Egelman, and David Wagner. 2019. Privacy Controls for Always-Listening Devices. In *Proceedings of the 2019 New Security Paradigms Workshop (NSPW 2019)*. ACM. https://doi.org/10.1145/3368860.3368867

[24] Moira McGregor and John C. Tang. 2017. More to Meetings: Challenges in Using Speech-Based Technology to Support Meetings. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 2208–2220.

[25] Donald McMillan. 2017. Implicit Interaction Through Machine Learning: Challenges in Design, Accountability, and Privacy. In *Internet Science*. Springer International Publishing, 352–358.

[26] Donald McMillan and Antoine Loriette. 2015. Living with Listening Services: Privacy and Control in IoT. In *Internet Science*. Springer International Publishing, 100–109.

[27] Donald McMillan, Antoine Loriette, and Barry Brown. 2015. Repurposing Conversation: Experiments with the Continuous Speech Stream. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 3953–3962.

[28] Pierrick Milhorat, Stephan Schlögl, Gérard Chollet, Jerome Boudy, Anna Esposito, and Gianni Pelosi. 2014. Building the next generation of personal digital assistants. In *2014 1st International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*. IEEE, 458–463.

[29] Todd F. Mozer and Pieter J. Verneulen. 2015. Background speech recognition assistant. US Patent 8,996,381.

[30] Pardis Emami Naeini, Sruti Bhagavatula, Hana Habib, Martin Degeling, Lujo Bauer, Lorrie Faith Cranor, and Norman Sadeh. 2017. Privacy Expectations and Preferences in an IoT World. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*. USENIX Association, Santa Clara, CA, 399–412. https://www.usenix.org/conference/soups2017/technical-sessions/presentation/naeini

[31] Helen Nissenbaum. 2004. Privacy as contextual integrity. *Wash. L. Rev.* 79 (2004), 119.

[32] Christi Olson and Kelli Kemery. 2019. Voice report: From answers to action: customer adoption of voice technology and digital assistants. https://advertiseonbing-blob.azureedge.net/blob/bingads/media/insight/whitepapers/2019/04%20apr/voice-report/bingads_2019_voicereport.pdf (last accessed on July 27, 2019).

[33] Antti Oulasvirta, Aurora Pihlajamaa, Jukka Perkiö, Debarshi Ray, Taneli Vähäkangas, Tero Hasu, Niklas Vainio, and Petri Myllymäki. 2012. Long-term Effects of Ubiquitous Surveillance in the Home. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp '12)*. ACM, New York, NY, USA, 41–50.

[34] Sarah Perez. 2019. Smart speakers hit critical mass in 2018. https://techcrunch.com/2018/12/28/smart-speakers-hit-critical-mass-in-2018/ (last accessed on August 5, 2019).

[35] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 640, 12 pages.

[36] Martin Porcheron, Joel E. Fischer, and Sarah Sharples. 2017. "Do Animals Have Accents?": Talking with Agents in Multi-Party Conversation. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 207–219.

[37] Matthew Purver, Thomas L. Griffiths, Konrad P. Körding, and Joshua B. Tenenbaum. 2006. Unsupervised Topic Modelling for Multi-party Spoken Discourse. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics (ACL-44)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 17–24.

[38] Yang Shi, Yang Wang, Ye Qi, John Chen, Xiaoyao Xu, and Kwan-Liu Ma. 2017. IdeaWall: Improving Creative Collaboration Through Combinatorial Visual Stimuli. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 594–603.

[39] John M. Simpson. 2017. Home Assistant Adopter Beware: Google, Amazon Digital Assistant Patents Reveal Plans for Mass Snooping. https://www.consumerwatchdog.org/privacy-technology/home-assistant-adopter-beware-google-amazon-digital-assistant-patents-reveal. Accessed: 2019-04-04.

[40] Stefan Sommer, Andreas Schieber, Kai Heinrich, and Andreas Hilbert. 2012. What is the Conversation About?: A Topic-Model-Based Approach for Analyzing Customer Sentiments in Twitter. *International Journal of Intelligent Information Technologies* 8 (01 2012), 10–25.

[41] Mark Sullivan. 2019. Supply-chain sources say Facebook Portal sales are 'very low'. https://www.fastcompany.com/90419009/supply-chain-sources-say-facebook-portal-sales-are-very-low.

[42] Alexandra Vtyurina and Adam Fourney. 2018. Exploring the Role of Conversational Cues in Guided Task Support with Virtual Assistants. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 208, 7 pages.

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

[44] Eric Zeng, Shrirang Mare, and Franziska Roesner. 2017. End User Security & Privacy Concerns with Smart Homes. In *Proceedings of the Thirteenth USENIX Conference on Usable Privacy and Security (SOUPS'17)*. USENIX Association, Berkeley, CA, USA, 65–80.

A SURVEY INSTRUMENT

A.1 The initial study instructions

In order to make sure that participants understood the study goal, we included a brief comprehension check after the consent form and study instructions. If a participant failed to respond correctly, we thanked them for their time and ended the survey. Below are the instructions and the comprehension check.

You're probably familiar with (and maybe even use) smart voice assistants, like Alexa, Siri, and Google Home. You address them by their "wake-word" (such as "Hey Alexa", "Siri", "Ok Google"), then ask a question or give a command.
For our survey, we ask you to imagine that there's a new voice assistant on the market: the Smart Speaker. Unlike today's devices, you don't need to say specific words to wake up the Smart Speaker, because it is always ready to help you. The Smart Speaker can also provide services and suggestions based on conversations you have with other members of your household.

Comprehension Check Question:

Which of the following is true of the device in this survey?

- The device is always on and can provide services and recommendations based on your current conversation without being explicitly invoked.
- The device reacts to your conversation only when explicitly invoked, for example when you say specific word, such as "Alexa", "Siri", "Ok Google."
- The device adds a video streaming feature and allows you to watch your favorite movies and shows on a big screen.
- The device is waterproof and can be used in a bathroom or swimming pool.

A.2 Survey questions

Each participant was presented with an audio recording of a conversation as well as the transcript, starting with the following introduction:

"Imagine the voice assistant recorded your conversation (as shown below). Please carefully listen to a 2-3 minute audio recording and think of the services that could be provided to you based on this conversation. There is also a transcription of the conversation for your convenience."

Sample conversation transcript:

A: I gotta clean up in here this place is just totally trashed cause I've done nothing this week but study and be sick. I've got a really bad dental problem. Or something with my mouth.

B: Poor Mom.
A: Think I've got a sinus infection or something. Don't touch the cookie..
B: Okay.
A: Please... I know it's tempting... What I'd like you to do is put those cans away please
B: Where? Where? Oh there they are...
A: Yeah... there they are.
B: A one... A two... Let's make the statue of hamburger city.
A: Mm
B: The statue of Coke.
A: Yeah.
B: The swinging barn
A: You're just a swinging kid Steve.
B: Yeah... You don't know the half of it.
A: I don't know the half of it... do I? Yeah... Oh man. Hey Steve why don't you give your iguana a little bit of banana too he'd probably really like some... He'd probably really like some banana.
B: Thanks Mom.

A: [laughter] Oh and I think this is Robbie's shirt and his uh Harley-Davidson scarf. Right?
B: Hmm?
A: Isn't that Robbie's shirt and uh Harley-Davidson scarf from this summer? I want to give that back to them tomorrow when we go over for his birthday.
B: I need to get Robbie a um present too.
A: Yeah... what do you think he'd like to have.
B: I'm not sure but we could go over to Toys 'R' Us.
A: It seems to me I brought the Toys 'R' Us catalog back with me.It's right over there.
B: Okay.
A: why don't you have a look at it and see if anything comes to mind for something you think Robbie would like to have for his birthday.
B: Well I have some things in here for Christmas
A: Yeah I know you probably see things in there that you want for Christmas but right now we're thinking about him and his birthday.
B: Okay.
A: And I gotta get started on this chicken pizza...

Then, the question about services followed:

The company behind the Smart Speaker wants to provide services based on the conversation you just heard. The services can be offered right away (with the Smart Speaker talking to the people in the conversation) or at a later point.
Think about the information in this conversation and come up with at least 3 features the Smart Speaker might offer to its users. Please describe these services in detail and give specific examples of how that service would assist the people in the conversation.

- Service 1: ...
- Service 2: ...
- Service 3: ...
- Add another service ...

Next, for each of the proposed services, we asked the following questions. The direction of scales (i.e., as presented below and with reversed order) was balanced across the sample.

(1) [Perceived level of comfort] How would you feel if the voice assistant used the conversation you just read to provide you the Service ...
   - Very uncomfortable (24.3%)
   - Somewhat uncomfortable (28.5%)
   - Neither comfortable nor uncomfortable (8.1%)
   - Somewhat comfortable (23.2%)
   - Very comfortable (15.8%)

(2) [Attention check question] Please select the option somewhat comfortable
   - Very uncomfortable
   - Somewhat uncomfortable
   - Neither comfortable nor uncomfortable
   - Somewhat comfortable
   - Very comfortable

(3) [Service usefulness] To what extent would you find Service ... useful or useless?

- Completely useless (4.2%)
- Slightly useless (7.2%)
- Neither useful nor useless (11.2%)
- Slightly useful (46.4%)
- Very useful (30.9%)

(4) [Behavioral intent] Which of the following options better describes the decision you would eventually take for service …
- I would not allow the Voice Assistant to use the conversation I just heard **at all** to provide me service …(47.16%)
- I would allow the Voice Assistant to use **the parts of the conversation I just heard, but not fully** to provide me service …(25.04%)
- I would allow the Voice Assistant to use the conversation I just heard **fully** to provide me services …(27.8%)

(5) [Reasons explaining the behavioral intent] Explain why would you make that decision

We also asked a conversation-specific question:

(6) [Recording sensitivity] How personal do you think the conversation in the audio you just heard is in general?
- Very personal (43.8%)
- Somewhat personal (30.3%)
- Neither personal nor ordinary (3.9%)
- Somewhat ordinary (14.6%)
- Very ordinary (7.3%)

We finished with the background and demographic questions:

(7) [Smart speaker owner] Do you have a voice assistant device (i.e. Alexa, Google Home) in your home?
- Yes (42.7%)
- No (57.3%)

(8) How do you identify your gender?
- Male
- Female
- Other: …
- Prefer not to say

(9) How old are you (in years)?

(10) What is the highest degree or level of school you have completed? (If you're currently enrolled in school, please indicate the highest degree you have received.)
- Did not complete high school
- High school or GED
- Some college/ Associate's degree
- Bachelor's degree
- Master's degree
- Doctoral degree