Farida Yeasmin

PRIVACY ANALYSIS OF VOICE USER INTERFACES

Master of Science Thesis
Information Technology and Communication Sciences (ITC)
Examiners: Prof. Mikko Valkama
Prof. Tom Bäckström
March 2020
ABSTRACT

Farida Yeasmin: Privacy Analysis of Voice User Interfaces
Master of Science Thesis
Tampere University
Information Technology
March 2020

A voice user interface (VUI) allows a user to interact with an application or system through voice or speech commands. A voice assistant device (VAD) primarily uses VUI to communicate with the user. The popularity of VADs is increasing due to voice is one of the most natural modes of human communication. The popularity of VADs and VUIs raises privacy concerns such as unwanted location tracking of a person using voice fingerprint data, unwanted recording of a private conversation, and unwanted habit detection of a person.

In this thesis, our research goal is to understand the emotional experiences and privacy expectations of users when they interact with VADs. We also want to identify suitable notification methods for human-to-VAD communication. In our research: (1) We categorize privacy in five types based on our background study, (2) We define four contexts based on location of the user and sight of view of the user to the VAD, (3) We analyze emotional experiences and privacy expectations of VADs based on a user study, and (4) We implement and evaluate a privacy-aware birthday pizza application that applies privacy-aware notification methods to inform users when collecting private data from conversations (e.g., birthdate).

Our evaluation shows that people have a strong preference for maintaining two types of privacy while interacting with VADs. These are location privacy and listening to private conversations. Our findings also revealed that users have several privacy expectations: 1) Consent for use of private data, 2) A feature to forget private data, 3) A function to turn off the device, and 4) Private data can be used only for positive purposes. We also find that users prefer visual and application-based notification when interacting with a VAD at home, while they prefer audio and application-based notification when at a classmate’s house.

Keywords: Privacy, Voice user interface, Notification modalities, Voice assistant device

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.
PREFACE

This master’s thesis is performed at Aalto university to fulfill the requirement of the Masters of Science degree in Information Technology of Tampere University.

Firstly, I would like to thank professor Tom Bäckström to allow me to work at his research group at Aalto University. Secondly, I would like to express sincere gratitude to my supervisors professor Mikko Valkama and professor Tom Bäckström for guiding the thesis work. I also thank my instructor, Sneha Das, for providing constant motivation, suggestions, and advice during my thesis. Finally, I would like to thank many others who have provided feedback to the thesis.

I thank my family members, especially, my husband who has always motivated me by his positive attitude and showed his immense patience for my work. Last but not least, this thesis is dedicated to my parents who supported me with unconditional love and understanding through all stages of my life.

Espoo, 15th March 2020

Farida Yeasmin
CONTENTS

1 Introduction ......................................................... 1
   1.1 Research problem and motivation .......................... 1
   1.2 Research methodology .................................... 3
   1.3 Main contribution and relevance of the work ............... 3
   1.4 Organization of the work .................................. 4

2 Background ......................................................... 5
   2.1 Voice user interface ....................................... 5
      2.1.1 Voice assistant devices ............................... 5
      2.1.2 Alexa voice service architecture .................... 6
   2.2 Privacy ....................................................... 7
      2.2.1 Why is privacy important? ............................ 8
      2.2.2 Contextual concept of privacy ......................... 9
      2.2.3 The GDPR and speech data ............................ 10
      2.2.4 Seven types of privacy ................................ 10
      2.2.5 Privacy impact assessment ............................ 11
      2.2.6 Privacy in data communication ......................... 12
   2.3 Privacy in voice assistant devices ......................... 13
      2.3.1 Gaussian mixture models for privacy-preserving speaker verification and identifications .................. 13
      2.3.2 The insecurity of home digital voice assistants-
            Amazon Alexa as a case study .......................... 14
      2.3.3 Privacy by design ..................................... 14
   2.4 User experience ............................................. 15
      2.4.1 UX model .............................................. 16

3 Research Methodology ............................................. 18
   3.1 Research question and contexts ........................... 18
      3.1.1 Privacy types and contexts .......................... 19
   3.2 Research methodology ...................................... 21
      3.2.1 Background study ..................................... 21
      3.2.2 Research design ....................................... 21
      3.2.3 Implementation ........................................ 22
      3.2.4 Evaluation of the implementation ..................... 22
   3.3 User study method .......................................... 22
   3.4 Data analysis method ....................................... 23
      3.4.1 Qualitative method .................................... 24
      3.4.2 Quantitative method ................................... 24
3.5 Summary ......................................................... 25
4 Implementation .................................................. 26
  4.1 Alexa skill ...................................................... 26
    4.1.1 Developing a skill with Alexa ......................... 27
  4.2 Features of birthday pizza application .................... 27
5 Results ........................................................... 30
  5.1 Emotional experiences and privacy expectations of VUI .... 30
    5.1.1 Privacy of the person — voice identification .......... 31
    5.1.2 Privacy of behavior and action — habits detection .... 32
    5.1.3 Privacy of location and space — location tracking ... 34
    5.1.4 Privacy of data and image — listening private conversation .. 34
    5.1.5 Privacy of thoughts and feelings — emotion detection .... 35
    5.1.6 Summary of emotional experiences and privacy expectations from VAD ........................................ 36
  5.2 Privacy notifications ........................................ 38
    5.2.1 Own house, looking at VAD ............................. 38
    5.2.2 Own house, not looking at VAD ......................... 38
    5.2.3 Classmate house, looking at VAD ....................... 39
    5.2.4 Classmate house, not looking at VAD .................. 39
    5.2.5 Summary of own house contexts ........................ 39
    5.2.6 Summary of classmate house contexts .................. 40
  5.3 Results from the implementation .......................... 40
  5.4 Summary ....................................................... 43
6 Discussion ....................................................... 44
  6.1 Evaluation of research questions and findings .............. 44
    6.1.1 Limitations .............................................. 45
  6.2 Evaluation of the implementation .......................... 45
  6.3 Future work ................................................. 46
7 Conclusion ....................................................... 47
References ......................................................... 48
Appendix A Story board and interview questionnaire ............. 51
  A.1 Story board .................................................. 52
  A.2 Context description and questionnaire ...................... 54
  A.3 Questionnaire for notification modalities .................. 59
  A.4 User study for implementation evaluation .................. 62
Appendix B Code ................................................... 66
  B.1 Interaction model ........................................... 66
  B.2 Background code ........................................... 69
LIST OF FIGURES

2.1 Alexa voice service workflow ....................................................... 7
2.2 PIA process for evaluation voice user interface ............................... 12

4.1 Privacy aware voice user interface for birthday pizza application ....... 29

5.1 Users at own house and looking at the device ................................. 39
5.2 Users at own house and not looking at the device ............................. 40
5.3 Users at classmates house, looking at the device ............................ 41
5.4 Users at classmates house, not looking at the device ....................... 42
**LIST OF TABLES**

2.1 Alexa’s visual modalities for notifications ........................................... 7

3.1 Contexts for emotional experiences and privacy expectations ................. 20

3.2 Context for identification of notification modalities .............................. 20

5.1 Emotional experiences and privacy expectations in voice identification .... 32

5.2 Emotional experiences and privacy expectations for habit detection (N = 8) 33

5.3 Emotional experiences and expectations for location tracking (N = 8) ....... 35

5.4 Emotional experiences and privacy expectations for listening to private conversations (N = 8) ................................................................. 36

5.5 Emotional experiences and expectations for emotion detection by VAD (N = 8) ................................................................. 37

5.6 Summary of emotional experiences ..................................................... 37

5.7 Results from prototype evaluation .................................................... 42
## LIST OF SYMBOLS AND ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASR</td>
<td>Automatic speech recognition</td>
</tr>
<tr>
<td>BLS</td>
<td>Boneh-lynn-shacham</td>
</tr>
<tr>
<td>DPIA</td>
<td>Data protection impact assessment</td>
</tr>
<tr>
<td>GDPR</td>
<td>General data protection regulation</td>
</tr>
<tr>
<td>HDVA</td>
<td>Home digital voice assistant</td>
</tr>
<tr>
<td>ISO</td>
<td>International organization for standardization</td>
</tr>
<tr>
<td>NLU</td>
<td>Natural language understanding</td>
</tr>
<tr>
<td>PIA</td>
<td>Privacy impact assessment</td>
</tr>
<tr>
<td>SSML</td>
<td>Speech synthesis markup language</td>
</tr>
<tr>
<td>UX</td>
<td>User experience</td>
</tr>
<tr>
<td>VAD</td>
<td>Voice assistant device</td>
</tr>
<tr>
<td>VUI</td>
<td>Voice user interface</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

A voice user interface (VUI) interacts with users using a spoken language. An application that primarily uses VUI as an interaction method is called a VUI application. A VUI application uses prompts, grammars, and dialog logic to interact with users (Cohen, Giangola and Balogh 2004). The prompts define speeches that are played by the VUI application during a dialog with the user. The grammars define a set of possible options a user can say in response to the prompt. The dialog logic establishes a set of actions that a VUI application can perform in response to the user’s speech. The VUI interaction method is unique (Cohen, Giangola and Balogh 2004) in several ways compared to other interaction methods. In the following, we list two uniqueness characteristics of VUI. First, in everyday life, humans interact with others using a spoken language. Through this, humans develop assumptions, conventions, and expectations for voice communication. Some of these assumptions are explicitly spoken during a conversation while others are implicit and expected that the other party understands. However, these assumptions may not hold for a VUI application when it communicates with the user. Second, humans do not always think about word choices and information shared when speaking with other humans. However, they may act differently when using other interaction media e.g., writing, or touch. The same may apply when the user talks to the VUI application.

The use of voice for interacting with different types of applications is growing rapidly. For example, most dominant mobile operating systems and hence mobile phones support VUI applications such as Google Assistant, and Apple Siri. Using these applications, a user can interact over voice with the mobile phone and applications within. Several smart home controller and speaker devices such as Amazon Echo and Google Home primarily use voice for interacting with the users. However, the rapid growth of such devices (Bohn 2019) and the unique nature of the voice interaction compared to others raise privacy concerns. For example, several privacy breaches (Coldewey 2019) related to smart devices listening to private conversations created a lack of confidence and confusion on the privacy of speech interfaces.

1.1 Research problem and motivation

A user may bring expectations and assumptions from human-to-human communication, which may lead to false expectations in voice interaction with a VUI based application. A user may also have an incorrect assumption about privacy (Wicker and Schrader 2010)
practices of a VUI based application, e.g., policies regarding types collected data and uses of the collected data. For instance, Amazon Echo may upload snippets of user conversations (Su 2019) with the device to a cloud-based service where actual humans analyze these conversations. A user may be unaware that his or her private conversations may be listened to by another human without explicit permission. These types of mismatch between the user expectations and practices by the VUI based applications can result in privacy breaches for a user. Moreover, a user may have different privacy expectations for different types of private data. For example, a user may expect location data to remain private while purchasing habits can become public with the consent of the user.

Traditionally, the privacy practices of an application or service mainly depend on the privacy policies and privacy notifications. However, recent researches (Jensen and Potts 2004, McDonald and Cranor 2008) show that privacy policies written in natural languages are infective and difficult to understand for a user. Thus, several earlier pieces of research (Kelley et al. 2009, Schaub, Balebako et al. 2015) propose summary notices as an alternative to a long privacy policy. However, there is still concern that most of these notices are not useful for the user, and eventually, the user will ignore those. These limitations bring the concept of context (Nissenbaum 2009), which plays a significant role in reducing the mismatch between the privacy expectation of a user and the privacy policies of an application. For example, a user can expect a health monitoring application may collect the user’s health-related data while a gaming application should not collect the user’s health-related data. Similarly, the privacy expectation of the user can vary based on socio-cultural background, personal trait, location of use, time, and technical know-how of the user. For example, a user may expect a strong form of privacy in a shared environment compared to a secluded environment. Based on this idea, Schaub et. al (Schaub, Könings and Weber 2015) propose short-form privacy notices depending on the user’s context, so that a user can make an informed privacy decision.

There are few earlier works related to privacy concerns of VUI. Moorthy et al. (Easwara Moorthy and Vu 2015) studied private and non-private information transmission over a VUI in the context of location. Their study revealed that users are more cautious when transmitting private information from a public place. However, their study is limited by the lack of granularity of privacy and context. Additionally, they do not focus on the user’s expected notification when communicating private information over a VUI.

A device that primarily uses a voice user interface for communication is called a voice assistant device (VAD). Currently, we use VAD for controlling smart homes, calling, and helping in our daily life. A VAD receives voice commands, performs an action based on the command and provides notification for its action to the user. Similar to other interfaces, a user interacts with the VAD using a voice interface and receives feedback either in voice, audio, or visual modalities. However, there is a concern over the effectiveness of the VAD devices to notify the actions of the device. Several recent reports (Coldewey 2019) revealed that major commercial VADs performed actions without a proper under-
standing of the user commands. There is also concern about the privacy of the VAD, such as 1) a VAD may fingerprint a person voice, 2) it can record conversations and commands from a private discussion, 3) it may able to detect emotions and behaviors from user conversations, 4) it may track a user location and share information with others. Moreover, privacy becomes a primary concern for these devices, as most of these devices are used in a shared environment.

There is a lack of research in the existing literature regarding privacy experiences, expectations, and privacy notification modalities (e.g., audio, visual) for VADs. In this thesis, we want to understand these concepts from the perspective of a VAD. Formally, our research questions are:

- What are the emotional experiences and privacy expectations of the user, when communicating with a VUI?
- How do the above experiences and preferences vary for changing context?
- What are suitable notification modalities for VAD and do the modalities change with the change of context?

1.2 Research methodology

We followed sequential exploratory research methodology (Creswell 1999; foodrisc 2019) to understand the user’s emotional experiences and privacy expectation for human to VUI communication. This research methodology provides a comprehensive and complete understanding of the research problem. We choose this method to identify the emotional experiences and privacy expectations for VUI As there is a lack of existing researches in this domain. The same approach is used to determine the suitable privacy notification modalities for VAD devices.

In both cases, we first understand the problem domain by performing a background study. Later on, we define a set of contexts for the study. We conduct user studies to collect data and use the data to evaluate our research question. We have identified the emotional experiences, privacy expectations, and suitable privacy notification modalities for VUI. We implemented a simple privacy notification system as an Amazon Alexa skill. We evaluate our implementation by performing a user study.

1.3 Main contribution and relevance of the work

The main contributions of this thesis are as follows:

- We have identified emotional experiences and privacy expectations for human-to-VUI communication. We also identified suitable notification modalities for VAD.
- We have evaluated the relationship between emotional experiences and privacy
expectations. We have assessed the relationship in several different contexts.

- We have implemented a privacy notification application for a VAD based on the result of our study. The application can notify users about storage private data during a conversation. We evaluated the privacy notices provided by the implemented application.
- We have critically analyzed our work and suggested future improvement areas.

1.4 Organization of the work

The organization of this thesis are as follows:

Chapter 2, Background discusses earlier pieces of research works related to voice user interfaces, privacy in data communication, privacy in user interfaces, and theory of usability study for privacy analysis.

Chapter 3, Research methodology presents our research questions, research methodology, and data analysis methods.

Chapter 4, Implementation provides a detailed workflow about our implemented privacy notification system of privacy for a voice assistant device.

Chapter 5, Results includes the results from data analysis for our defined research goals.

Chapter 6, Discussion analyzes our findings in several aspects, e.g., based on earlier works, limitations of the study, and future work.

Chapter 7, Conclusion presents the overall summary of the thesis.
2 BACKGROUND

This section explains the concepts and related technological tools used in this thesis.

2.1 Voice user interface

A voice user interface (VUI) allows a user to interact with voice-enabled applications through speech commands. VUI uses speech recognition to perceive spoken commands and provide a response to the commands. The VUI converts speech-to-text for user commands and text-to-speech for the reply. The elements of a VUI includes grammars, dialogs, and prompts. During conversations with the user, the synthesized speech (the prompts or system messages) is played to the user. To each prompt of the system, the possible response of the user is defined by the grammar where the system can only understand the included grammars (e.g., words, sentences, or phrases). The system response is based on the defined dialog logic of the system. The following example is an interaction between a caller and a car rent information application.

System: Hello, Welcome to the car rent shop...Which brand of car do you want to rent?
Caller: I don’t know.
System: Okay! No problem! What is your budget to rent a car?
Caller: Around five hundred euro

In this application, the system prompts are pre-recorded by a voice actor. The system listens to the caller following the prompt “Which brand of car, do you want to rent?”. The system then uses grammar to accommodates inputs from the caller such as “don’t know”, “no” or “yes”. The system uses the dialog logic for the next step. In this example, the prompt for the budget of the car is based on the dialog logic. If the dialog logic succeeds, the system will provide information to rent a car to the caller.

2.1.1 Voice assistant devices

Voice assistant devices (VAD) primarily use voice as an interaction modality with the user. These devices include an intelligent assistant that can perform tasks for a user based on commands. The intelligent assistants in the VAD continuously listen to a keyword as a
wake-up word. When it hears the keyword, it records the voice of the user and sends it to a specialized server for interpretation and process the command (Hoy 2018). The server replies to the intelligent assistant with appropriate responses, e.g., execute the user requested command or request for further information. VAD devices can perform a range of services. For example, provide information for weather, set an alarm, make a to-do list, and read an audiobook. There are several commercially available VAD devices in the market — for example, Amazon Echo, Google Home Assistant, and Apple HomePod.

2.1.2 Alexa voice service architecture

At the beginning of the user interaction with the device, the user gives a voice command to the Amazon Alexa device. The device uses an on-device buffer, which is continuously overwritten to detect the wake word inside the room with the help of acoustic patterns. After detecting the wake word, e.g., "Alexa," the device streams the voice command to the Alexa cloud. The device does not stream any voice responses to the cloud until the action button on the device is pressed, or the dedicated wake word has not been detected. The streamed voice command is first received by the automatic speech recognition (ASR) system. The ASR receives the voice command and converts the speech into a text string. Later on, the text string is sent to the natural language understanding (NLU) system. This system work interprets the recognition result, and based on the outcome, NLU produces an intent.

We present Alexa voice service architecture in Figure 2.1. In the figure, the possible intent is a weather intent. Once NLU interprets the possible intent, it then forwards the intent name to the proper skill application. The skill application then formulates its response by taking the raw data, and speech synthesis markup language (SSML) construct in a textual format. To generate the audio response, the text-to-speech system takes the produced SSML and translates the skill output to speech. The audio or voice response is then streamed to the appropriate device from where the user gets the voice responses for the provided voice command.

Alexa system stores several types of data inside the Alexa data store. As described above, it stores the user’s raw voice command data before voice commands are sent to the ASR. Each voice recordings processed through ASR and NLU are also stored in the Alexa data store to improve the Alexa service later on.

Alexa Echo devices provide a range of built-in notification both in visual and audio format to inform users about performed actions. Table 2.1 lists built-in notifications from Alexa Echo devices. Note that, for critical operations, e.g., wake up word, Alexa Echo devices use both audio and visual notifications.
Table 2.1. Alexa’s visual modalities for notifications

<table>
<thead>
<tr>
<th>Task</th>
<th>Visual action</th>
<th>Audio action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wake word: e.g., “Alexa”</td>
<td>Round circle light becomes blue</td>
<td>Provide a sound and wakes up</td>
</tr>
<tr>
<td>Continuous listening</td>
<td>The light stays blue</td>
<td>No sound</td>
</tr>
<tr>
<td>Call is active</td>
<td>Light is on spinning green mode</td>
<td>No sound</td>
</tr>
<tr>
<td>Incoming call</td>
<td>Lights color changed to pulsing green</td>
<td>Provide a continuous sound till the user speak to alexa</td>
</tr>
<tr>
<td>Muted microphone: Alexa does not re-</td>
<td>Red lights</td>
<td>Provide a sound while give any command in this state 2) Provide spoken words to inform about the state.</td>
</tr>
<tr>
<td>sponse to command</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do not disturb mode</td>
<td>Circle color is flashing purple</td>
<td>Provide a sound while give any command in this state . Provide spoken words to inform about the state.</td>
</tr>
<tr>
<td>Neutral mode</td>
<td>No color</td>
<td>No sound</td>
</tr>
</tbody>
</table>

2.2 Privacy

In the engineering of networking systems defining privacy has proven to be a critical topic. The author Wicker et. al (Wicker and Schrader 2010), defined privacy from the perspectives of “The right to privacy” and “Privacy as contextual integrity”. The author of the “The right to privacy”, (Brandeis and Warren 1890) attempts to connect the privacy of new technologies and potential dangers of it. In various common law precedents,
the author identified the tort of preexisting privacy. The author proposed the right to be let alone as the right of privacy, which implies that an individual can be able to restrain their personal information from being published in a public forum. Their concept of privacy, right to be let alone, can be generalized as a metaphor for a zone of seclusion. Here the zone is an area of personal information that is controlled by an agent. The notion of safety and secrecy of agent lies in part of a zone. We can view a zone of seclusion as privacy where personal information of an individual can have access control to own information. For instance, without fear of manipulation or censure, an individual can develop and experiment and lead an autonomous life. We can also define privacy as a combination of several variables. These variable are: 1) Nature of the situation and contained context; 2) The information of the nature related to that context; 3) The receiving agents roles; 4) Relationships of information subjects; 5) The subject shares the information on what terms and further of the terms dissemination.

2.2.1 Why is privacy important?

We summarize importance of privacy based on the article (Solove 2014).

- **Limit on power:** Privacy is a limit on different sectors of power, such as a limit on government power and private sector company’s power. The more power someone can have over us, the more they know about our data since, to make crucial decisions in our personal lives, the personal data is used. It can also be used to affect our reputations; influence our decisions; make a tool to control us; can cause us great harm if the data is in the wrong hand.

- **Respects for individuals:** It is disrespectful if an individual’s desires of wish to keep something private are ignored without a constrain reason to do so. Privacy is about respecting individuals. However, sometimes individual desire to keep something private may not succeed because of other values conflicts with privacy. In many cases, the privacy of an individual does not cause major injury to other individuals.

- **Trust:** Breach of trust is a breach of the confidentiality of an individual. For instance, to maintain a professional relationship trust is the key. Such as our relationships with a lawyer or doctors rely on trust. Similarly, we trust other parties as well.

- **Freedom of thoughts and speech:** One of the important parts of privacy is to have freedom of thoughts and speech. Privacy of freedom of thoughts provides a person to explore ideas from outside of the mainstream by reading or watching everything which their family or friends may dislikes. Privacy also provides freedom of speaking where an individual can have the freedom to speak on an unpopular message to share their own opinion or criticism for other messages with the world.

---

1 A tort is an act that injures a person in a way that the injured person may take legal action against the wrongdoer. In legal terms, torts are civil wrongs.
• **Ability to change and have second chances:** Privacy nurtures the ability to change and grow and provides a second chance to move beyond the mistakes which are performed by us in the past. We should not expect that all the mistakes are shielded; however, some errors should be hidden from others to facilitate growth, foster, and development.

• **No obligation:** If every time we have to ponder how we will explain or justify each activity to others, it can be a severe obligation to oneself. Thus privacy gives right to the people, so that they do not need to justify to others.

Privacy is also essential for freedom of social and political activities, control over one's life, maintaining appropriate social boundaries and reputation management.

### 2.2.2 Contextual concept of privacy

Hull et. al (Hull, Lipford and Latulipe 2011) present Nissenbaum's contextual integrity framework in the context of social networking and how two particular social phenomena blogging and webcam change equation of privacy.

In every sphere of life, information flow is different. There are no context-specific norms that control information flows. For example, an individual can share his or her child's medical report with a doctor, but it is not appropriate to share the report with everybody. According to Nissenbaum, this information sharing is called norms of distribution. Norms of distribution are one of the types of two norms provided by Nissenbaum. On the other hand, another norm is called appropriateness. For example, it is appropriate to share a report with a doctor, but it is not appropriate to share religious information with employers unless an individual gives permission to share the information publicly or in a very limited circumstance. Thus, Nissenbaum underlined two points, such as information is not context-free rather obstructed, and there is no universal privacy norm. These privacy issues have theoretical implications that lead to pursuing three principles. These are 1) Individual's privacy protection from government agents, 2) Limit the access of sensitive information, and 3) Removing unauthorized access from private information. However, with these three principles, it is difficult to capture intuitive privacy issues. The author proposed that the above principles applied to Facebook by creating multiple groups and by customizing the group's accessibility.

Again, Nissenbaum explains that social networking, such as blogging, has privacy issues. An individual's information can be shared with everybody, which can break the individual's privacy and that individual can face a negative impact on his or her life, which is not expected. Webcamming is another social networking system where anyone can spy on anyone else and broadcast the information with the whole world, which also related to Nissenbaum norms. On Facebook, it is possible to make many friends, which is not possible to make it offline. It seems that users of Facebook share their private information on Facebook, which is accessible by the specific Facebook friends. This causes privacy
breach due to the disclosure of private information. Norms of appropriateness suggest that a user's Facebook friend cannot be disclosed to other friends with the same personal information of the user. Again, norms of distribution suggest that one may share his or her friend's picture on Facebook but cannot share the information with the friend's employers. Thus Facebook has privacy issues the same as observed for webcamming and blogging. Finally, the paper suggests ensuring the user's privacy so that users feel secure while using a social networking system. User's personal information cannot be shared with others unless they give permission.

2.2.3 The GDPR and speech data

General data protection regulation (GDPR) does not provide a concrete privacy policy law for speech data and speech technology as the legal and technical communities do not provide a general regulation in which cases the privacy should apply. The paper (Nautsch et al. 2019) aimed to provide some possible regulation in speech privacy for both legal and technical perspectives. In the legal privacy part, privacy in speech biometric data and how sensitive data can be identified, and rules for using sensitive data have been discussed. In the technical privacy perspective, how privacy is characterized in terms of regulation has been discussed with a clearer picture.

GDPR inaugurated biometric data policy, which is personal data. Personal data includes physical, physiological, or behavioral characteristics that help to characterize a person identification, e.g., identification of facial images. Biometric data is a kind of sensitive data as it has personal identification of a person. Sensitive data also includes other information such as health data, political opinions, and religious beliefs. One cannot publish another person's sensitive data unless it is permitted to be. This regulation can be applied in speech data privacy. An example of the scenario is that audio or video recordings can be found on the Internet, but it doesn't mean that the speech data of that audio or video can be used by others. To reduce the risk of personal data processing, legal experts need to perform a Data protection impact assessment (DPIA).

2.2.4 Seven types of privacy

This section explains seven types of privacy (Finn, Wright and Friedewald 2013). The privacy includes 1) the privacy of the person, 2) the privacy of behavior and action, 3) the privacy of personal communication, 4) the privacy of data and image, 5) the privacy of thoughts and feelings, 6) the privacy of location and space and 7) the privacy of association (including group privacy).

The privacy of the person deals with personal body characteristics such as genetic codes and biometrics. This privacy is comprised of an individual's feelings of freedom. As
privacy deals with body characteristics, thus we consider voice fingerprint as the privacy of the person. Voice fingerprint is identifiable by a voice assistant device.

The privacy of behavior and action deals with people’s sensitive issues such as sexual preferences, habits, political activities, and religious practices. This privacy provides the freedom to behave in a public or private space without being monitored by others. In our research, habit is considered as a notion of privacy of behavior and action, where we hypothesized that the user’s habit is identifiable by a VAD.

The privacy of communication deals to avoid communications being intercepted. It includes such as mail interception, interception of wireless communication, or recordings. The ability to communicate without being interrupted or recorded with the help of different media gives benefits of a free discussion for a wide range of views and options, and it also enables the communications sector growth.

The concerns of privacy of data and images include an individual’s private data. An individual's data should not be available to other parties, and an individual can have control over use of that data. The privacy provides individual confidence of empowerment. We consider listening to the private conversation of an individual by a VAD as a concern of privacy.

Privacy of thoughts and feelings includes an individual’s right for not sharing thoughts and feelings with others. We again hypothesized that a VAD can identify an individual’s feelings based on that voice tone and thoughts can be shared with the device intentionally or unintentionally by a person.

Privacy of location and space provides the right to move in public or semi-public spaces freely without being tracked. A VAD can track the location of a person using voice fingerprint data from one area to another. Again, indoor tracking can happen when the user talks to other devices. In such cases, the device can identify the distance through variation of voice frequency.

Finally, privacy of association gives people freedom to choose about their association without being monitored. The privacy also includes group privacy.

2.2.5 Privacy impact assessment

Privacy impact assessment (PIA) is a systematic process to identify and manage potential privacy invasiveness or impacts on privacy (Clarke 2009). The PIA has the following main characteristics: 1) A PIA is anticipatory in nature, i.e. it is performed advance to understand the privacy impact 2) A PIA has a broad scope which can contain all types of privacy 3) A PIA address both the problem and the solutions to them. Traditionally, it is performed by an organization to review privacy impacts from projects, products, processes, and work flows.
PIA process

We outline a simplified process in Figure 2.2 for PIA based on existing literature study (Friedewald et al. 2016; Bieker et al. 2016). Our process follows an iterative three steps: 1) Preparation stage 2) Evaluation stage and 3) Report and safeguard stage.

2.2.6 Privacy in data communication

This section discusses about privacy of data communication based on several earlier studies.

A survey of security and privacy in cloud computing

The author of the paper (Zhou et al. 2010) investigates the privacy and security concerns of cloud computing systems. The privacy and security concerns are investigated
in the perspective of different applications and infrastructures. The author provided five concerns such as, confidentiality, availability, data integrity, control, and audit about security issues as they find out that provided security concerns by the companies are not adequate enough. The goal of availability is that users can use the system from any suitable places of the users. One example can be the virtual machines, which can provide individual users on-demand services.

A privacy-preserving public auditing technique for data storage security in cloud computing

The author of the paper (Wang et al. 2010) provides a scheme for a privacy-preserving public auditing technique for data storage security in cloud computing. In this technique, a public key based homomorphic authenticator and random masking are used, so that third-party auditors cannot collect data content of the user while auditing the user’s file on the cloud server. Specifically, Boneh-lynn- shacham (BLS) based signature is used for multi-task auditing.

2.3 Privacy in voice assistant devices

Privacy is a significant concern for voice assistant users. These devices listen all the time so that they can respond to users’ commands. Several VAD companies like Apple, Google, Amazon, and Microsoft insist that unless the user does not speak the wake word to the VAD, the assistant does not start a recording of voice (Hoy 2018). However, there are multiple cases where a voice assistant device was recording all the time and sending those recordings back to the VAD’s server (Tung 2017). Even if we assume that the voice assistant device manufacturers are careful and scrupulous when developing their assistant devices, there is a potential for data to be leaked, stolen, or used to accuse users. In another incident of a murder investigation, authorities issue a warrant to a company to retrieves the VAD of the suspect (Buhr 2016).

2.3.1 Gaussian mixture models for privacy-preserving speaker verification and identifications

A person’s voice and manner of speaking or speech is one possible way to identify an individual's characteristics. The authors of the article (Hull, Lipford and Latulipe 2011), developed a framework to protect privacy for speaker verification and identification. This method uses a secure multi-party computation protocol and it operates on encrypted speech. The paper proposes a client-server model in which a user runs a client program that uses public and private key cryptography to protect the speaker’s voice. The
public key is shared with others while the private key is only known to the user. This approach eliminates speaker identification and possible phishing attacks during speaker verification. This approach requires that the system can only store encrypted speech data provided by the enrolled users. Therefore, a secure multi-party computation protocol can be used for speaker identification and verification to achieve privacy in the speech interface.

2.3.2 The insecurity of home digital voice assistants - Amazon Alexa as a case study

Home digital voice assistant (HDVA) like Amazon Alexa and Google home are now getting popular among people due to their ability to perform actions based on voice commands such as checking the weather, and making an order. The author of the paper (Lei et al. 2017) investigates Alexa's security problem in access control. Alexa depends on single-factor authentication and it takes voice commands while no people are around the device. Thus, the author suggested another physical authentication factor, i.e., presence-based access control. Their paper selects the Alexa device as Alexa offers a huge range of voice service skills (more than 10,000), which are far more than other voice assistant devices, e.g., Google Home. Later, the findings can be implemented on other HDVA devices, as well.

The insecurity of HDVA due to the frankness of the voice channel may lead users to security threats. The solution to the addressed problem is straightforward. Before taking a voice command of users, the voice command shall be authenticated by their voice biometrics. Though the remedy is not natural as a user's voice may vary due to different scenarios such as illness. Note that even with improvement, the user’s voice can be replayed by an attacker.

For user authentication, the article (Feng, Fawaz and Shin 2017) proposes to use wearable devices that collect the vibration of the signals and continuously match the collected signals with the received signals by HDVA devices. The proposed remedy may not be suitable for the users due to the accessibility of the devices at home. The author of paper (Lei et al. 2017) finds out that acoustics attacks are launched while users are absent in the house. Thus, the author of the article proposes and validates the idea of a Virtual Security Button on the HDVA devices, which will help the device to distinguish indoor motions from no motion and outdoor motions.

2.3.3 Privacy by design

Privacy by design is a framework which defines a set of rules to design UX that complies with user’s expectation for privacy. These design rules guide designers for UX design,
which complies with existing privacy laws and respects the privacy of the user. Wicker et al. (Wicker and Schrader 2010) presents five rules for privacy-aware design. These are:

- **Full disclosure of data collection**: Disclosure is a long-recognized approach used when collecting user data. However, in many cases, disclosure policies become so complicated that a user fails to understand the true meaning of disclosure. Wicker et al. (Wicker and Schrader 2010) propose a couple of criteria for well-designed disclosure. First, the description of the disclosure should include the types of collected data at a granular level, how long the data will be stored, and the purpose of data collection. Second, the collected data should fall under the privacy policy, and that policy should be clear to the user. Third, the privacy policy and the disclosure text should be intelligible to the user. A long legal privacy text can easily distract the user from protecting privacy. A recent study (McDonald and Cranor 2008) shows that an average user would require 244 hours to read privacy policies for sites he or she visits in a year. Thus, a context-based small notification is preferred over long privacy policy disclosure text (Friedman 2019). The preferred approach here is "Just-in-time" notification and explain why the data collection is required. The notification can point to a link for the overall privacy policy.

- **Consent for data collection**: The consent is an agreement to disclose private data. The guiding rule is the user should explicitly acknowledge any use of private data by communication technology tools. For example, the user should click an agreement on-screen before using a service. The other approach for receiving consent could be written form or voice confirmation. The EU GDPR directive (Voigt and Von dem Bussche 2017) states that the collector should receive explicit consent from the data subject before usage.

- **Minimize collection of personal data**: The guiding principle is the collection of personal data is required for functionalities of the technology. Private data collection for training and testing purposes should be avoided.

- **Minimize identification of personal data with individuals**: The collected personal data should be anonymized when possible. If the collected data is stored, the data should be stored in an anonymized format. If there is a need for usage of personal data for individual identification, there can be a separate processor that links anonymized data to an individual.

- **Minimize data retention**: Data should be deleted if not required for the functionality of technology. If data is required for the future function of the technology, the data should be protected with sufficient technical means.

### 2.4 User experience

User experience is emotions and responses that a person shows when using or going to use a particular product, service or system (DIS 2010). UX fulfills more than pragmatic
needs, i.e., it acknowledges the subjective, situated, complex, and dynamic encounters of a person. According to Hassenzahl (Hassenzahl and Tractinsky 2006) UX is a result of an internal state of a user, the characteristics of a designed system, and the context. The definition of user experience from the international organization for standardization (ISO) (DIS 2010) indicates a holistic aspect of UX. This emphasizes the temporal aspects relevant to the earlier anticipations of the usage concurrently, the experience based on the usage.

In the following list, we provide a few characteristics (Rezaei 2015) of user experience.

- First, UX illustrates the experimental aspects, such as temporal aspects and multiple facets of UX. It includes preferences, emotions, beliefs, behaviors, physical and psychological responses, and accomplishments that happen in different time periods such as before, during, and after use.
- Second, UX is a result of the presentation, the system’s interactive behavior, the image of brand, functionality, the user’s physical and internal state that is the result of the prior experiences, skills, personality, attitudes and the context of the use.
- Third, UX illustrates the role of usability, which is in association with UX. Usability can be interpreted from users’ personal goals, which can include emotional and perceptual aspects.

To conclude, the definition of this UX focuses on four aspects. These are experimental aspects, i.e., users perceptions and responses that result from the use or going to use of a product, service or system; experience of temporal aspects (user’s knowledge of the usage of product and system in different periods such as before, during and after usage); influencing aspects (all affecting factors of UX consists of the features of an interactive system, the characteristics of the user, and context of use), and finally, as system attributes usability may affect user experience.

### 2.4.1 UX model

In user experience, we need UX model (Kaipainen 2019) to

- Understand the two main attributes of UX, pragmatic and hedonic. Pragmatic attributes refer to, with products or services, how users are able to complete tasks. On the other hand, how much users enjoy using the products or services is called hedonic attributes.
- Guide user experience design and evaluation. In the UX design and evaluation, the UX method helps to identify which components to consider in UX design and to choose the appropriate methods to measure user UX.

One of the prominent models of UX is Hassenzahl's model. The model assumes that when users use a product or service, they assign some attributes while using the product.
The attributes can be grouped into four categories: identification, manipulation, stimulation, and evocation. These attributes can be categorized into two higher-level attributes: pragmatic and hedonic. The pragmatic attributes (manipulation) relate to the product's practical usages and functions, and hedonic attributes (identification, stimulation, and evocation) refer to the psychological well-being of users (Twobenches 2019).
3 RESEARCH METHODOLOGY

The research aims to understand privacy experiences and expectations for human-to-voice user interface (VUI) communication where the VUI is provided by a voice assistant device (VAD). Additionally, we want to identify suitable modalities to inform privacy notification by VAD devices. For this, we first perform exploratory research to understand privacy expectations and determine the appropriate modalities for human-to-VUI communication. After that, we implement a prototype based on user privacy expectations and further evaluate our implementation by a user study. In this section, we explain the research questions, research methodology, and data analysis methods for the overall research scope of this thesis.

3.1 Research question and contexts

Our research questions are the following: When a user communicates with a VAD using VUI (1) What are the emotional experiences and privacy expectations of the user? (2) How do these privacy experiences and expectations differ for the change of context? (3) What are suitable notification modalities for VAD, and how modalities vary with the change of context?

Now, we will explain each part of our research questions. The first part of our research is identifying emotional experiences and privacy expectations. Human emotion is a critical element of information privacy. Several earlier researches (Stark 2016; Mozilla 2019) places emotion as a critical part of privacy discussion. Privacy is a subjective, sensory, and contextual event at a personal level. Privacy depends on our feelings, preferences, and our deeds. These traits define our subjective feeling of privacy in a given condition. In this thesis, as a first step, we would like to understand emotional experiences for human-to-VUI communication. The second part of this research question is privacy expectations. A user has certain privacy expectations when communicating with VUI. These are defined as desired privacy expectations. For example, a user may expect that online service will not store data without consent or, by default, would not share data with others. In this thesis, as a second step, we have identified desired privacy expectations for human-to-VUI communication.

The second part of the research is understanding the impact of context for emotional experiences and expectation for privacy. The context, such as location, groups, and
individual traits, can impact privacy experiences. For example, a user's privacy expectation for VUI in the home environment is different compared to the public environment. The user may not want any discussion regarding sensitive information in the open environment.

The final part of our research is to identify suitable privacy notification modalities for VAD devices. Currently, written privacy policies act as a primary tool for notifying the user about the privacy of a service. Several earlier researches (Jensen and Potts 2004; McDonald and Cranor 2008) have described that a lengthy privacy policy is time-consuming to read, difficult to understand, and therefore ignored by the users. Several earlier works (Kelley et al. 2009) propose summary privacy notices in visual format as an alternative to lengthy privacy policy statements. However, there is a lack of prior research regarding what are suitable notification modalities for privacy notices by VAD devices during voice communication. In this thesis, we want to identify suitable notification modalities for privacy notification by VAD devices.

### 3.1.1 Privacy types and contexts

This section explains data practices and scenarios for our research. Earlier research by Reidenberg et. al (Reidenberg et al. 2015) shows that people are more concerned about unauthorized collection, sharing, and retention of personal data. In another research, Finn et. al (Finn, Wright and Friedewald 2013) categories privacy in seven types based on the data type. From these seven categories, we select five types of privacy to analyze the privacy impact of VAD. These are (1) Privacy of the person, e.g., voice fingerprint (2) Privacy of behavior, e.g., habits (3) Privacy of location and space, e.g., location (4) Privacy of data and image, e.g., private conversation, and (5) Privacy of thoughts and feelings. Note that the remaining two, i.e. the privacy of communication and the privacy of association, are not particularly relevant for VAD. The privacy of communication, although a valid privacy type, is a general privacy problem for any interception in digital communication. So, it is not specific to VAD. Similarly, the privacy of association applies more to humans rather than to devices.

We define two high-level contexts for human-to-VUI communication based on the location of the user. In the first context, the user is located at home and speaking with their VAD device. In the second context, the user is located at the classmate's house and speaking with a guest device. Similarly, we define two comparable contexts for human-to-human communication. These are (1) the user is speaking with their relatives and (2) the user is speaking with a friend of a classmate. For human-to-human communication, we assume that a VUI device located in own home can be considered as a relative. Similarly, a VUI device located at friend's home can be considered as a friend of the classmate. Our assumption is based on the trust we place on devices and persons for these two contexts. We also assume that the friend of a classmate is not previously known to the user. We use the privacy experiences from the human-to-human scenario as a comparison guide
for human-to-VUI communication. Table 3.1 shows the scenarios in tabular format.

**Table 3.1. Contexts for emotional experiences and privacy expectations**

<table>
<thead>
<tr>
<th>Example of privacy types (See section 2.2.4)</th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Voice identification (e.g., voice fingerprint)</td>
<td>home alone, own device</td>
</tr>
<tr>
<td>2) Habits</td>
<td></td>
</tr>
<tr>
<td>3) Location and space</td>
<td></td>
</tr>
<tr>
<td>4) Private conversations</td>
<td></td>
</tr>
<tr>
<td>5) Emotions</td>
<td></td>
</tr>
<tr>
<td>1) Voice Fingerprint</td>
<td>Classmate’s house, guest device</td>
</tr>
<tr>
<td>2) Habits</td>
<td></td>
</tr>
<tr>
<td>3) Location and space</td>
<td></td>
</tr>
<tr>
<td>4) Private conversation</td>
<td></td>
</tr>
<tr>
<td>5) Emotions</td>
<td></td>
</tr>
</tbody>
</table>

To identify suitable notification modalities for VAD devices, we use the same two contexts used for the identification of privacy experiences of VUI. These are 1) the user is located at home alone (own device), and 2) the user is located at the classmate’s house (guest device). We further granulate each of these contexts using visual orientation perspectives, i.e. 1) the user is directly looking at the VAD while communicating with it, 2) the user not looking at the VAD when interacting with it. We asses the most suitable notification modalities based on these contexts. Table 3.2 shows the context for identification of privacy notification modalities.

**Table 3.2. Context for identification of notification modalities**

<table>
<thead>
<tr>
<th>Example of privacy types (See section 2.2.4)</th>
<th>Context</th>
<th>Visual orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Voice identification (e.g., voice fingerprint)</td>
<td>Home alone and own device</td>
<td>Looking or Not looking</td>
</tr>
<tr>
<td>2) Habits</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3) Location and space</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4) Private conversation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5) Emotions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1) Voice Fingerprint</td>
<td>Classmate’s house and guest device</td>
<td>Looking or Not looking</td>
</tr>
<tr>
<td>2) Habits</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3) Location and space</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4) Private conversation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5) Emotions</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.2 Research methodology

We follow a sequential exploratory research (Creswell 1999; foodrisc 2019) methodology in this thesis. We choose this method as there is a lack of existing researches that identify privacy implication for human-to-VUI communication. Additionally, this research methodology can provide a more comprehensive and complete understanding of the research problem than either qualitative or quantitative approaches alone. We use the following steps in our research.

3.2.1 Background study

We studied about privacy concerns for several other technology areas such as mobile devices and web-based interfaces. We study psychological aspects of privacy to understand the need for privacy for VUI. We also explore different types of notification modalities to notify users about privacy events. Additionally, we select two primary contexts for evaluating the emotional experiences and privacy expectations of VUI (see Section 3.1.1).

3.2.2 Research design

The research process follows an iterative step. In each level, we use background study, user study, and evaluation to validate the findings from the earlier stage. The data collection goals are explained below:

- We first collect qualitative data to understand users emotional experiences and expectations of privacy for human-to-VUI communication. Next, we understand how people emotionally react and expect privacy in their interactions with other people. The aim is to provide users a clear sense of understanding of the impacts of privacy in human-to-human communication. Finally, we compare the user’s emotional experiences and expectations between the human-to-human and the human-to-VUI(device) setting.

- We collect quantitative data to identify the suitable notification modalities for human-to-VUI communication. We also relate appropriate notification modalities with the emotional experiences.

- We analyzed and evaluated the collected qualitative and quantitative data from the user study (see data analysis in Section Result 5) and derive the relation between the two.
3.2.3 Implementation

Based on the findings from the user study, we have implemented an application that provides privacy-aware notification for human-to-VUI communication. We have applied the findings from user study and followed privacy by design guidelines for the developed application. The application works on Alexa powered voice assistant devices.

3.2.4 Evaluation of the implementation

We have conducted a second user study to evaluate the implemented application. We have assessed the application by performing a user study that calculates the effectiveness based on psychometric (Rust and Golombok 2014) response of a user.

3.3 User study method

The section explains the methodology used for the user study. We followed a couple of necessary steps from the Nielsen Norman model (Norman 2005) for performing the user study. These are (1) Defining the goals, (2) Defining format and setting of the study, (3) Participants recruitment, (4) Participant’s task to match the goals of the study, (5) Pilot study and improvement of the task, (6) Conduct the study, and (7) Analyzing data and derive result.

**Goal:** This goal is to understand the emotional experiences and privacy expectation of voice user interface. We also want to identify preferred notification modalities in voice communication.

**Study format and setting:** We decided to perform two semi-structured interviews. The first interview targets the exploration phase, and the second one focuses on the evaluation of the implementation. We conduct the study, face-to-face in a lab setting, where we choose a moderate study process which guides participants through the study and helped them to understand the questions. Another reason for the moderate study was to ensure that we get better insights into the participants’ concerns for privacy and have opportunities to ask clarification of their answers. We assumed that the device located in a home environment, i.e. either in their own home or classmate house. Multiple users may use the device. There can be more than one device.

**Participants:** In the exploratory phase, we recruited eight participants using a random process from a known set of groups, which consists of university students who know about the latest trend in technology. The participants are from technical fields. This ensures that they know voice user interface and voice assistant devices. All participants are aged above eighteen years. The participants are recruited through a social media group by advertising an interview description. They were selected in the first come first
An interview with a participant took approximately 1 hour and 30 minutes to complete, and participants are rewarded with two movie tickets for participating in the interview. The interview pattern is the same for all the participants. In the implementation evaluation phase, we recruited five participants and followed a similar procedure.

**Participant's tasks:** In the exploratory phase, the participants have two specific tasks: (1) Answer to several open-ended questionnaires to collect qualitative data, and (2) Multiple-choice questions to collect quantitative data.

In the implementation evaluation phase, the participants evaluate our implementation using a Likert scale (Albaum 1997). The participants evaluate two implementation notification features for storing personal data. The Likert scale is a rating scale for measuring attitudes and opinions from the participants. Using the Likert scale, we asked the participant's agreement or disagreement for several dimensions of the implementation. We used the Likert scale on a 5-point ordinal scale to measure each construct reflectively. Likert scale: 1 = Totally disagree (TD); 2 = Disagree (D); 3 = Neither agree nor disagree; 4 = Agree (A); 5 = Totally agree (TA).

**Pilot study:** We perform a pilot study with three participants. We use the result of the pilot study to improve the scenarios and questionnaires.

**Conduct study:** We conduct the exploratory interview in two days. The study format was semi-structured. During the study, first, we educate the users on how VUI and VAD work using a storyboard. Second, we describe scenarios, one by one, and ask open-ended questions regarding the user's emotional state for each type of privacy. The answer to the open-ended questions is recorded, which we manually transcribed to written format in a later stage. Finally, we ask questions with a close-ended multiple-choice answer. The participants recorded the answer to the multiple-choice questions in the printed paper.

We interview evaluation of the implementation in a single day span. Unlike earlier, this study format is structured. We ask the user to rate our privacy-aware notification application.

**Analyzing data and result:** We use the qualitative data analysis method for open-ended questions. We use a quantitative data analysis method for multiple-choice questions. We explain the data analysis methods in Section 3.4.

### 3.4 Data analysis method

In the following, we describe the data analysis methods for both qualitative and quantitative data.
3.4.1 Qualitative method

We use thematic analysis (Caulfield 2019) to analyze the qualitative data. This is a method for analyzing qualitative data, which is applied to a set of interview texts to analyze and identify common themes within data. It is a typical method to find out people's views, opinions, knowledge, and experiences from a set of qualitative data. We use the following four steps based on this method.

- **Familiarization and codify the data**: We transcribed from the recorded audio to the textual format. From the text version, we codify the topics by reading it several times. In this context, codify means, first, highlighting the relevant sentences from the transcribed audio and then coming up with shorthand labels or code.

- **Generating themes and reviewing themes**: We combine several codes into a single theme by manually and systematically. We then compare the generated themes with the actual data set and verify the presence of the theme in the data set. We iterate the process again for producing a better theme.

- **Defining and naming the themes**: In this step, we define and label the themes. Defining a theme includes formulating the exact meaning of each theme and identify how it helps us to understand the data. Naming themes provides a concise and easily understandable name for each theme.

- **Emotional analysis for Themes**: We use Plutchik’s wheel of emotions (Plutchik 2001; Gomes 2017) as a topic for theme analysis. The Plutchik emotion wheel defines four primary emotions (or emotional category): joy, trust, fear, and surprise. These emotions have four polar opposites: sadness, disgust, anger, and anticipation. The classification of themes into an emotional category based on Plutchik let us better understand the user’s stated emotion and reduces the ambiguity.

Note that, we performed the analysis for both human-to-human communication and human-to-VUI communication.

3.4.2 Quantitative method

We performed a quantitative analysis both in the exploratory and evaluation of the implementation phase. In the exploratory phase, we collect quantitative data for suitable notification modalities. These data use a nominal scale. We follow a cross-tabulation approach for this analysis. This analysis provides an easy way to correlate between study variables. We use the following two steps for analysis:

- **Define measurement scale**: We define notification modalities as a variable of the scale.

- **Connect descriptive statistics with the data**: We link the frequency, i.e., number of times a notification modality is stated by the participants.
For the evaluation of the implementation, we collect data based on a Likert scale. Each participant rates two implemented privacy notification features in several dimensions, e.g., helpfulness, requirement, and understandability. Here, we use the following approach for data analysis.

- **Define measurement scale:** We use the dimension of each question as a variable of the scale.

- **Connect descriptive statistics with the data:** We link the mean, i.e., the average rating for each dimension to the scale.

### 3.5 Summary

This section describes the research problem of this thesis. We explain the reason for selecting the five privacy types and define the contexts of our study. We also discuss the research methodology and provide the reason for selecting the research methodology.
4 IMPLEMENTATION

We implemented a privacy-aware Birthday pizza skill (application) for Amazon Alexa (Amazon 2019a). The Birthday pizza application logic is written in Python programming language, while the interface part is written using JavaScript Object Notation (JSON). We use the Alexa skill kit, which is a collection of APIs, tools, and documentation for the Birthday pizza application. We also use Amazon developer console \(^1\) for development, build, deploy, and test for the skill. The Amazon developer console supports both Python and NodeJS to write application logic. We choose Python because it allows rapid prototyping and a better-debugging facility.

A user can invoke the Birthday pizza skill by using a voice command to the Amazon Alexa. The skill is divided into two segments: frontend and backend. The frontend includes an interaction model (JSON schema), which defines the pattern to interact with the user. The Birthday pizza skill model has two custom Intents: (1) CaptureBirthdayIntent (2) CapturePizzaIntent. The CaptureBirthdayIntent has two slots, namely month and day, which store the user spoken birthday. The CapturePizzaIntent has one slot to store the user spoken pizza type. The skill model also includes a set of Alexa provided default Intents for Cancel, Help, and Stop activity. The backend includes the logic for processing user requests. We use standard Python packages to write the backend logic. The only exceptions are Alexa provided ask_sdk_core and ask_sdk_model \(^2\) libraries which are required to interact with the model file. The backend consists of Intent handlers to process user-requested intent. The developed skill is deployed in the Amazon developer console simulator. We have tested the deployed application, both using the Alexa simulator and using the Amazon Echo Dot device.

4.1 Alexa skill

Alexa is a voice service provided by Amazon. An Alexa skill is similar to an application for Alexa, similar to a mobile application for a mobile operating system. One of the significant characteristics of Alexa’s skill is that it is a primarily voiced driven application. An Alexa skill listens to voice commands and respond when necessary. An Alexa skill can be deployed and published via Amazon Alexa market place. A published application can be used from Alexa enabled devices, e.g., Amazon Echo Dot.


\(^2\) https://github.com/alexa/alexa-skills-kit-sdk-for-python
Alexa skill uses voice-first interaction modality (Amazon 2019b) compared to the traditional mouse, keyboard, or touch-based interaction modalities. Humans tend to speak in a non or semi-structured way during the voice interaction. Due to this, VUI based application adapts to various voice commands to understand the actual meaning and intent from the conversation. For this, Alexa makes available a natural language processing engine for all applications.

4.1.1 Developing a skill with Alexa

An Alexa skill consists of two segments: frontend and backend. The frontend consists of an interaction model which is a collection of intents, utterances, and dialog. An intent represents an action that the user can perform with the skill. A set of intents form the core functionality for a skill. The utterances are sample words that a user says to call the intents. The dialog model defines the information required to run the skill and prompts to confirm information during a conversation with the user. The interaction model includes a mapping between the intents and the utterances. Each skill has a name that identified as invocationname. The user speaks the invocationname to start the application. The frontend also contains a mapping to the backend that receives intents and acts upon those intents.

The backend contains application-specific logic to handle intents. The logic is written as an AWS Lambda Function (Amazon 2019c). In the AWS Lambda function, each incoming requests by the frontend and returns an appropriate response to the frontend. The Lambda function should include a handler for each intent defined in the interaction model so that it can process requests for the specified intent. The returned response is spoken to the user using text to speech tool of Alexa.

4.2 Features of birthday pizza application

The application takes a pizza order from the user for an upcoming birthday. The application can also store the birthday for delivering pizza for future birthdays. We show the Birthday application workflow in Figure 4.1. In the Figure, the shorthand A represents Alexa, while the shorthand U represents the user.

In this thesis, we have identified several privacy expectations of VUI, which are explained in the Chapter 5. We have identified that users prefer audio-based notification and confirmation when the user is communicating with the classmate’s VAD, i.e. a guest device. Thus, we develop an application that provides audio notifications and confirmation dialog before storing privacy-sensitive data. Additionally, the application uses privacy-aware design principles defined in Section 2.3.3. The application uses audio-based small notifications to inform and take confirmation from the user before storing the birthday of the user. The notification follows best practices for data collection disclosure design. The
notification design details are provided below:

(1) Can I Save your (2) birthday as .... for (3) future birthday orders. We (4) do not share birthdays with other applications and you can in future (5) remove birthday when requested.

Here, (1) Can I Save and (2) birthday refer to the data that requires user confirmation. This option provides an opt-in choice for the user to store data or not. The (3) future birthday orders defines the purpose of storing the data. The (4) do not share and (5) remove text informs user about the privacy policy of the application.

In voice interaction, the user often does not think when providing confirmation. Due to this, our application also uses distinct audible two short Beep sound to guide users that sensitive data has been stored. For this, we use the Speech Synthesis Markup Language (SSML) (Amazon 2019d) feature of Alexa.
Figure 4.1. Privacy aware voice user interface for birthday pizza application
5 RESULTS

In this chapter, we present results of our research work. We divide the result into two sections. In the first section, we present results for emotional experiences and privacy expectations in several defined contexts. We also present our findings for suitable notification modalities in VAD. In the second section, we offer an evaluation of the implemented application.

5.1 Emotional experiences and privacy expectations of VUI

We have performed semi-structured interviews to understand emotional experiences and privacy expectation for a human-to-VUI communication. We ask an open-ended question in the interview to collect qualitative data for five privacy types in two different contexts. Our selected privacy types are: (1) Privacy of the person, e.g., voice identification or fingerprint; (2) Privacy of behavior and action, e.g., habits; (3) Privacy of location and space, e.g., location data; (4) Privacy of data and image, e.g., private conversation; and (5) Privacy of thoughts and feelings, e.g., emotion. Our selected two contexts are: (1) The user is alone in his or her own home, and the user owns the VAD device; and (2) The user is at classmate’s house, and the classmate owns the VAD device. For each of these contexts, we have a comparable scenario for human-to-human communication. These are: (1) The user is speaking with relatives, and (2) The user is speaking with a friend of a classmate. In the second case, the user does not previously know the friend of the classmate. We analyze emotional experiences and expectations in four steps.

- First, we group collected emotional experience data into themes and presented those in a tabular form. We then group these themes into emotional categories based on Plutchik’s wheel of emotions (Plutchik 2001; Gomes 2017). Note that Plutchik’s wheel defines four primary emotions: joy, trust, fear, and surprise. These emotions have four polar opposites: sadness, disgust, anger, and anticipation. Second, we list the privacy expectations of the user.
- We provide a comparison for emotional experiences between a user who is communicating with the VAD device of a classmate and a user who is communicating with a friend of a classmate.
- We provide a comparison for emotional experiences between a user who is communicating with their VAD and a user who is interacting with relatives.
• We provide a comparison for emotional experiences between users’ own devices and classmate’s devices.
• Finally, we analyze privacy expectations for each context.

The questionnaire for data collection is attached in Appendix A. We collected data from eight (N = 8) participants. In the text, P refers to participants, e.g., P1 is participant number one. For all tables in this section, we use ordinal numbers to represent the total number of response that falls into the same privacy experience theme.

5.1.1 Privacy of the person — voice identification

In Table 5.1, we present emotional experience themes, categorization of these themes, and privacy expectations for both human-to-VAD and human-to-human communication. Below, we provide a list of comparisons between two contexts based on emotional experience.

• Classmate’s VAD vs. Classmate’s friend: The majority of the participants feel fear (creepy) about voice identification by a classmate’s friend. For example, a quote from one of the participants is, “It is creepy. My question is how he or she knows about me. If a random person knows about me, then it is more weird” (P7: Classmates friend). Similarly, almost the same number of participants feel in a stronger emotional form (e.g., Shocking) when the device performs the same. Thus, fear can be labeled as the most dominant emotional experience when an unknown (Classmate’s device) device performs voice identification.

• Own VAD vs. Relatives: The majority of the participants trust their device for voice identification. One participant state that “I would be not so shocked. Because This is my voice assistant, I bought it so I know it can recognize me” (P8: Own device). In contrast, only a few participants trust their relatives when relatives can identify the user’s voice. As a summary, we label trust as the dominant emotional experience for voice identification by their own devices.

• Own VAD vs. Classmate’s VAD: Participants show positive emotions (e.g., trust) for voice identification by own VAD while participants show negative emotions for voice identification by classmate’s VAD (e.g., Fear, Surprise). However, in both cases, participants are concerned about the security risk of voice identification. For example, one can use the voice identity of another.

Privacy Expectation: Participants expect that their own devices should able to recognize their voices. If the device is already used several times, then a classmate’s device can also identify their voice. For such cases, participants are not concerned about their privacy of voice identification. However, they expect that a VAD should not identify voice if a user has not interacted with it before. This means participants do not like sharing voice identification data from one VAD to another. The participants also expect consent
for voice identification and functionality to remove voice identification data from the device or service.

**Table 5.1. Emotional experiences and privacy expectations in voice identification**

<table>
<thead>
<tr>
<th>Category</th>
<th>Emotional experience themes</th>
<th>Human-to-VAD</th>
<th>Human-to-Human</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Classmate’s</td>
<td>Own</td>
</tr>
<tr>
<td></td>
<td></td>
<td>VAD</td>
<td>VAD</td>
</tr>
<tr>
<td>Trust</td>
<td>Flattering</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Common, normal, fine</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Ok, if i have talked several times</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Fear</td>
<td>Shocking</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fear of storing data</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Feel watched</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Creepy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surprise</td>
<td>Curious about how the device works or person know about it</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Surprising</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Disgust</td>
<td>Unsafe</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Not okay sharing with random people or device</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Security risk</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sadness</td>
<td>Unhappy</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Expectation</td>
<td>Ask to forget the fingerprint</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Require consent for fingerprint use</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Others can not access my fingerprint other than me</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.1.2 Privacy of behavior and action — habits detection

We present emotional experience themes, categorization of these themes, and privacy expectations of habit detection in both human-to-VAD and human-to-human contexts in table 5.2. Below, we provide a list of comparisons between contexts based on emotional experiences.

- **Classmate’s VAD vs. Classmate’s friend:** Participant’s strongly stated fear emotion when a classmate’s VAD collects data for habit detection. The reaction is almost similar for human-to-human communication, i.e., when an unknown person is iden-
tifying a person’s habit. However, some participants are positive about behavior
detection by VAD devices. One participant states "First of all, I think that it has a
good algorithm design that it is able all the things which I have ordered before and
it has stored the things in it so that I won’t have to repeat all the order again" (P4: Classmates device). Overall, we conclude that participants show fears for habit identification by VAD devices.

- **Own VAD vs. Relatives**: Participants feel relaxed or satisfying when their device can identify their habits. However, participants are not so positive when relatives identify the habits. In both cases, participants do not want their habit data to be shared with others. One participant stated that "It is okay for me. But, I won’t like it if the device shares my personal habits with others" (P6: Own device).

- **Own VAD vs. Classmate’s VAD**: Participants show positive reactions for their own device. Participants think that it is efficient if the device knows about their personal ordering habits. However, participants show negative emotions toward classmate’s devices for habit detection.

**Privacy Expectations**: Participants show curiosity about habit detection, i.e., how the device identifies habits. This means that participants want to know how the habit detection algorithm works and what data is stored for this process. Participants expect that the device should not share habit detection data with others. They also want to disable this feature if required.

**Table 5.2. Emotional experiences and privacy expectations for habit detection (N = 8)**

<table>
<thead>
<tr>
<th>Category</th>
<th>Emotional experience themes</th>
<th>Human-to-VAD</th>
<th>Human-to-Human</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Classmates</td>
<td>Own VAD</td>
</tr>
<tr>
<td>Joy</td>
<td>Relaxed</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>satisfying</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Trust</td>
<td>Data can be shared</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Fear</td>
<td>Fear of storing data</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Threatening</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Creepy</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Surprise</td>
<td>Curious</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Sad</td>
<td>Unhappy</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Expectation</td>
<td>Turn off the device</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Do not share data with</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>random people or device</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5.1.3 Privacy of location and space — location tracking

In Table 5.3, we present emotional experience themes, categories, and expectations for both human-to-VAD and human-to-human contexts. Below, we provide a list of comparisons between contexts based on emotional experience for the privacy of location and space.

- **Classmate’s VAD vs. Classmate’s friend:** Participants strongly demonstrate angry emotion when classmate’s devices track the location of the speaker. The emotions are similar for human-to-human communication when a friend of a classmate does the same. Note that, we grouped ‘Stop using the device’ and ‘Not acceptable’ topic as angry emotion as both of these topics present strong objection to the practice. One participant stated that “More disturbing. Very very uncomfortable. I will go to the police. Devices can do something because they are unconscious, but a human can not do it”(P1: Classmates friend).

- **Own VAD vs. Relatives:** Similar to the previous observation, participants stated angry emotion if their own device or their relatives do the same.

- **Own VAD vs. Classmate’s VAD:** In both cases, participants strongly stated that location tracking is a serious breach of privacy, be it by their own device or by the device of someone else. Participants show either distrust or angry emotions when devices track location.

**Privacy Expectations:** Participants expectation for privacy is that location data collected by the VAD can be used only when appropriate consent is provided. This is true for both the classmate’s device and own device. For example, one participant stated that “If I had allowed it then it will be fine, but again it should ask permission before doing it”(P3: Own device).

5.1.4 Privacy of data and image — listening private conversation

We present emotional experience themes, categories, and expectations for both human-to-VAD and human-to-human contexts in Table 5.4. Below, we provide a list of comparisons between contexts based on emotional experiences.

- **Classmate’s VAD vs. Classmate’s friend:** Participants strongly oppose (by showing angry emotion) when a classmate’s device listens to private conversations. One participant mentioned that if the device listens to their private conversation, they will press any button of the device so that it can not listen to their conversations. The emotions are similar for human-to-human communication when a friend of a classmate does the same.

- **Own VAD vs. Relatives:** Similar to the previous observation, participants feel
Table 5.3. Emotional experiences and expectations for location tracking (N = 8)

<table>
<thead>
<tr>
<th>Category</th>
<th>Emotional experience themes</th>
<th>Human-to-VAD</th>
<th>Human-to-Human</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Classmates VAD</td>
<td>Own VAD</td>
</tr>
<tr>
<td>Trust</td>
<td>Data can be shared</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Distrust</td>
<td>Security risk</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Violation of the privacy</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Data can not be shared</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Sad</td>
<td>Unhappy</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Angry</td>
<td>Stop using the device</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Angry</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Not acceptable</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Expectation</td>
<td>Require consent for data use</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>

angry emotions if their own device or their relatives do the same.

- **Own VAD vs. Classmate’s VAD:** For all scenarios, participants primarily express distrust and angry emotions when a device listens to private conversations. One participant stated that “I do not think it is okay. It does not matter who’s voice assistant device it is. It is not cool” (P5: Own device).

**Privacy Expectations:** Participants want a higher degree of control over any type of private conversation listened by devices. According to the participants, listening to private conversations and storing the data is not acceptable. Participants expect that the device should ask for consent from the user before listening on private conversations. They also expect that the collected data should not be shared with other devices. The participants also want responsible usage of collected data.

5.1.5 Privacy of thoughts and feelings — emotion detection

In Table 5.5, we present emotional experience themes, categories, and expectations for both human-to-VAD and human-to-human contexts. Below, we provide a list of comparisons between contexts based on emotional experiences.

- **Classmate’s VAD vs. Classmate’s friend:** Participants express negative emotion when a friend of a classmate or the device of a classmate try to detect emotion. However, forty percent of the participants think it is okay for a classmate’s device to identify the emotion.

- **Own VAD vs. Relatives:** In contrast to the previous observation, participants are more positive about detecting emotion by their own devices and relatives. Participants mentioned that their own relatives (especially parents and close relatives)
Table 5.4. Emotional experiences and privacy expectations for listening to private conversations (N = 8)

<table>
<thead>
<tr>
<th>Category</th>
<th>Emotional experience themes</th>
<th>Human-to-VAD</th>
<th>Human-to-Human</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Classmates VAD</td>
<td>Own VAD</td>
</tr>
<tr>
<td>Fear</td>
<td>Shocking</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fear of storing data</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Security risk</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Surprise</td>
<td>Confusing</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Curious about how it collect data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distrust</td>
<td>Violation of the privacy</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Disturbing</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Angry</td>
<td>Turn off the device</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Angry</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Not acceptable</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Expectation</td>
<td>Require consent for data use</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sharing insensitive privacy data is ok</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Not okay if used for negative purpose</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

should identify their emotions. One participant stated that “I would expect my relatives to ask me what happened to me” (P8: own relatives). An equal number of participants also express negative emotions when their own device detects emotion.

- **Own VAD vs. Classmate’s VAD:** Participants express mixed reactions when a VAD detects emotion, be it own or a classmates’ device. Some participants think it is positive, while others feel it breach their privacy. From this, we can summarize that the privacy of emotion is less important than other types of privacy.

Privacy Expectations: Participants expect collected data should only be used for positive purposes. Participants also hope that collected data will not be shared with others.

5.1.6 Summary of emotional experiences and privacy expectations from VAD

We present the dominant emotional experience for five different privacy types in Table 5.6. We conclude that participants are more concerned about privacy associated with the
Table 5.5. Emotional experiences and expectations for emotion detection by VAD (N = 8)

<table>
<thead>
<tr>
<th>Category</th>
<th>Emotional experience themes</th>
<th>Human-to-VAD</th>
<th>Human-to-Human</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Classmates VAD</td>
<td>Own VAD</td>
</tr>
<tr>
<td>Trust</td>
<td>Common or normal</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Fear</td>
<td>Fear of storing data</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Surprise</td>
<td>Curious about how the device works</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Surprising</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distrust</td>
<td>Violation of the privacy</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Angry</td>
<td>Stop using the device</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Not acceptable</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Angry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expectation</td>
<td>Not okay if information is used for negative purpose</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Unlike if the data is shared</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Use information for positive purpose is ok</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

location and listening to a private conversation. On the other hand, they are less concerned about emotion detection. The emotional experiences between human-to-human and human-to-VAD follow almost symmetrical patterns except for few cases, e.g., participants trust their devices to detect habit while they do not feel the same for their relatives when they identify their habits.

Table 5.6. Summary of emotional experiences

<table>
<thead>
<tr>
<th>Privacy category</th>
<th>Human-to-VAD</th>
<th>Human-to-Human</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Classmates VAD</td>
<td>Own VAD</td>
</tr>
<tr>
<td>Voice Identification</td>
<td>Fear</td>
<td>Trust</td>
</tr>
<tr>
<td>Habits detection</td>
<td>Fear</td>
<td>Trust</td>
</tr>
<tr>
<td>Location tracking</td>
<td>Angry</td>
<td>Angry</td>
</tr>
<tr>
<td>Listening to private conversation</td>
<td>Angry</td>
<td>Angry</td>
</tr>
<tr>
<td>Emotion detection</td>
<td>inconclusive</td>
<td>inconclusive</td>
</tr>
</tbody>
</table>

In the following, we list the main privacy expectations for VADs:

1. Require consent to use private data
2. Feature to forget private data
3. Feature to turn off the device
4. Data can be used for positive purposes

5.2 Privacy notifications

We have conducted a quantitative analysis to identify expected notification modalities for five types of privacy. We use the following steps for the analysis. First, we have explained the purpose of the study to the participants. Second, we asked the participants to select notification modalities from a multiple-choice type questionnaire. We provide seven options for each question. These options are: Audio (Distinguishable tone) / Visual (Lightbar) / SMS / E-mail / App / I do not want any notification / Others (specify). Participants can select one or multiple options for each question.

We design questions based on four contexts. These are: (a) Own house and looking at VAD; (b) Own house and not looking at VAD; (c) Classmate’s house and looking at VAD; and (d) Classmate’s house and not looking at VAD. We attach the questionnaires in Section A.

5.2.1 Own house, looking at VAD

When participants are at their own home and interacting with the device while looking at it, they preferred visual notification modality for emotion detection, listening to private conversations, and voice detection. For location tracking and habit detection, users prefer notification via app and audio, respectively. Figure 5.1 shows the result of our analysis. Overall, the user prefers visual notification for the home scenario.

5.2.2 Own house, not looking at VAD

When users are at their home and interacting with the device while not looking at it, they prefer app-based notification for habit detection and location tracking. Users prefer audio notification for listening to private conversations and visual notification for emotion detection. Overall, notification via application is preferred by users for most privacy types except for listening to a private conversation where users want audio-based notification. The audio-based notification has the benefit of informing users in an obtrusive way compared to other types. Figure 5.2 shows the results for this context.
5.2.3 **Classmate house, looking at VAD**

Audio is the most preferred notification for voice identification, listening to private conversations, and emotion detection when a user is at a classmate’s house and interacting with the VAD while looking at it. On the other hand, the application is preferred for location tracking and habit detection. Figure 5.3 presents the results for this scenario.

5.2.4 **Classmate house, not looking at VAD**

In the context of interacting with the device while not looking at it, users prefer application based notification for location tracking and listening to private conversations. Users prefer no notification for voice identification, habit detection, and emotion detection. Figure 5.4 shows the result of our analysis for this category.

5.2.5 **Summary of own house contexts**

As a summary, users mostly prefer visual notifications when they are looking at the devices. While they mostly prefer application based notifications when not looking at the devices. As an exception to the above rule, users prefer audio-based notifications for
5.2.6 Summary of classmate house contexts

As a summary, users mostly prefer audio notifications when they are looking at VADs located at classmates' house. While they equally prefer application or audio-based notifications when not looking at the devices. Users' preference for application based notifications imply that users want control over what to do with the notifications. There is also an equal preference for no notification when VADs perform voice, habit, and emotion detections. This implies that these three types of privacy are less important for the participants.

5.3 Results from the implementation

We implemented an Amazon Alexa skill by incorporating the privacy expectations from our study in Section 5.1.6. The skill uses audio modalities (derived from our study result) for privacy notifications. We implemented two types of audio notifications when storing private data, i.e., birthday. The first type is: (1) Audio-based explicit notification when a
private data is stored and (2) Distinct audio sound to notify users that a privacy-sensitive operation has taken place. We perform a user study to evaluate these two types of audio notifications. We evaluated each notification in four dimensions: 1) Helpful, 2) Noticeable, 3) Required, and 4) Sufficient. We evaluate each dimension using a 5-point Likert-scale, i.e. (1) Totally disagree; (2) Disagree; (3) Neutral; (4) Agree; and (5) Totally agree.

Our study goal is to understand the effectiveness of the audio modalities for privacy notifications. For this, we perform the user study with $N = 5$ participants. Participants answered a short questionnaire about their agreement or disagreement towards privacy-aware notifications after interacting with our prototype. Table 5.7 lists the participant's feedback regarding implemented privacy notifications. Participants overall rated a median of 4.5, which is in between agree and totally agree scale. Most of the participants totally agree that the notifications (both the audio confirmation and beep audio) provided by the VAD are helpful and noticeable to draw users' attention when a VAD stores personal data. All participants agree that audio confirmation and beep audio notifications are required to identify the privacy actions of the voice assistant devices. Similarly, participants believe these two notifications are sufficient for privacy actions.

Participants provided the following suggestions for the overall improvement of the application. They provided comments regarding the notification system. People expect that if a VAD stores private data, the stored data can be checked later on. One participant mentioned, "store it in the application, so it could be later checked." Participants also suggest
**Figure 5.4.** Users at classmates house, not looking at the device

**Table 5.7.** Results from prototype evaluation

<table>
<thead>
<tr>
<th>Participant Number</th>
<th>Audio confirmation is helpful</th>
<th>Beep audio is helpful</th>
<th>Audio confirmation is clear to understand</th>
<th>Beep audio is noticeable</th>
<th>Audio confirmation and beep audio is required</th>
<th>Audio confirmation and beep audio is sufficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td><strong>4.6</strong></td>
<td><strong>4.4</strong></td>
<td><strong>4.6</strong></td>
<td><strong>4.6</strong></td>
<td><strong>4</strong></td>
<td><strong>3.8</strong></td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td><strong>5</strong></td>
<td><strong>4</strong></td>
<td><strong>5</strong></td>
<td><strong>5</strong></td>
<td><strong>4</strong></td>
<td><strong>4</strong></td>
</tr>
</tbody>
</table>

that the beeping audio should happen before asking to store private data. That way, the user would be more focused on what Alexa asks, knowing that the next thing that he or she speaks is important. People also showed concern about security risks. They want the notifications to specify if the saved personal data is only used for the intended service or by other services as well.
5.4 Summary

This section summarizes emotional experiences and privacy expectations for human-to-VAD communication. We list and evaluate these experiences and expectations for five different types of privacy in two primary contexts: own home and classmate house. From the evaluation, we identify that location and listening to a private conversation are more privacy-invasive compared to other types of privacy. We also identify that users mostly prefer visual and application-based privacy notifications for devices in their home. While, in the classmates’ house, users mostly prefer audio and application-based privacy notifications. Furthermore, we implemented a prototype to validate the audio-based privacy notification. The evaluation shows that our implemented audio and beep notifications are helpful, noticeable, and required to draw attention of the user.
6 DISCUSSION

We provide a discussion of our thesis work in this chapter. We also provide a comparison of our work with earlier studies. In the end, we suggest a set of future works that will better explore privacy relationships for VAD devices.

6.1 Evaluation of research questions and findings

In this thesis, we evaluated privacy experiences and expectations of VADs for five types of privacy in two main contexts, i.e. own home and classmate’s house. We also identify the user’s expected notification modalities for each type of privacy in four different contexts.

We have identified that participants are more concerned about the privacy of location and listening to private conversations by VAD devices. They showed angry reactions when these types of privacy breaches are performed by VADs. On the other hand, they are less concerned about emotion detection by VADs. We also identified that participants expect the following from VADs: (1) Consent for use of private data, (2) A feature to forget private data, (3) A feature to turn off the device, and (4) Private data can be used for positive purposes.

We identify that participants mostly prefer visual notifications when they are looking at VADs at their own homes. While they mostly prefer application-based notifications when not looking at the devices at their own homes. In a classmate’s house setting, participants mostly prefer audio notifications when they are looking at the devices. While they equally prefer application or audio-based notifications when not looking at the devices.

We derived these findings using a sequential exploratory research method, i.e. a combination of qualitative and quantitative data collected through semi-structured interviews, and defined questionnaires. The open-ended interview questions let us uncover participant’s privacy experiences and expectations from VAD devices. We choose this method as there is a lack of research for the privacy requirements of VAD devices. We derive expected notification modalities for each privacy type from quantitative data collected using a defined set of questionnaires. Here, we guide the user to select the best notifications modalities among a set of modalities. We choose this method as we already know about a set of notification modalities supported by the VAD devices.

We argue that privacy is a complicated subject for VAD devices. There is a lack of ex-
isting research in this domain. We also argue that our research provides a set of basic
guidelines for privacy experiences and expectations from VAD devices. The findings in
this work can be taken into consideration to address the privacy aspects of VAD.

6.1.1 Limitations

Our study is performed in a lab setting with a limited number of participants. Additionally,
we primarily focus on the transcribed text to identify the emotions of the user. In some
cases, we detect the emotion from the tone of the user. However, we did not critically
relate to the behavior and non-verbal expressions with spoken words during interviews.
This may limit the outcome of our study.

We have used Plutchik’s wheel of emotions to categorize and analyze the emotional
experiences of a user. The Plutchik’s wheel of emotions defines four primary emotions:
joy, trust, fear, and surprise. We use this model for all types of privacy. We think an
additional analysis method could provide more granular findings, and is left for future
investigations.

6.2 Evaluation of the implementation

We developed an audio-based privacy notification for voice assistant devices, more specif-
ically a privacy-aware Birthday Pizza skill for Amazon Alexa. The application uses audio-
based notification when private data, e.g., the birthday, is collected and stored. For the
notification, we take findings from privacy expectations (Section 5.1.6) of VAD and privacy
by design approach (Section 2.3.3). These two states that the notification text should in-
form the user about 1) Target of data collection, 2) Opt-in choice for the user, and 3) user
control over the data. Designing a notification in such a way clarifies the goal of data col-
lection to the user. We also implemented an additional BEEP notification so that the user
can distinguish between privacy confirmation and normal confirmation, e.g., purchase
order.

We only implemented notification when the device collects private data from conversa-
tions. We did not implement privacy notifications for other types of privacy, e.g., voice
identification, habit detection, and location sharing. In the future, we plan to extend the
implementation to other types of privacy in future work. Note that we implemented the
notification within an application. This limits the usage of this notification method by other
applications. In the future, we plan to implement this as a library, so other applications
can easily incorporate the library.

We did not perform any tests to check the performance for applications using such notifi-
cations. We plan to analyze the performance implications of notifications in the future.
6.3 Future work

We use the semi-structured interview in a lab environment to gather data for privacy experiences and expected notification modalities for VAD. Data collected using the semi-structured in lab settings may not reflect the actual pattern compared to realistic field studies with behavior tracking. In a lab setting, participants may not be express their genuine emotions for different privacy scenarios. We have observed during the interview that a proportion of participants (25 percent) did not illustrate the answers for the asked questions. In the future, we plan to gather more illustrative interview by observing their behavior in a realistic field setting. For example, a participant may not be able to express emotions through answering questions. However, a person’s physical expression can be helpful to understand the emotions.

We only analyzed emotional experiences and expectations based on a defined five types of privacy in two primary contexts. Our defined five types of privacy, in some scenarios, may become coarse and fails to differentiate emotional experiences. For example, we analyzed the emotional reactions of participants when a VAD device tracks their habit. However, we did not further granularize different types of habits, e.g., purchasing habit and movie watch list habits. Thus, our study lacks granularity for analyzing emotional reactions.

From the user study, we gathered and analyzed quantitative data about the user’s expected notification modalities for five types of privacy. We did not present the notification in a real setting by a VAD during the user study (except during implementation evaluation). We assumed that participants of the user study could imagine the context and the presented notification modalities. Thus, it is likely that participants were unable to think which notification is suitable for each type of privacy. Additionally, our study may not provide adequate notification modalities options to the user. In the future, we plan to implement each notification modalities using a VAD and perform user study in target environments. The same approach can be used for analyzing privacy experiences.
7 CONCLUSION

The usage of VUI in voice assistant devices is gaining popularity as voice is one of the most natural modalities for human. The demand for VUI can become a serious privacy concern for many users. This situation can happen due to a user of a VUI may have false expectations and assumptions about privacy based on existing human-to-human communications. In this thesis, we uncover emotional experiences and privacy expectations for interactions between human-to-VAD. By conducting user studies and analysis of data, we identify that users strongly prefer maintaining the privacy of their location data and listening to private conversations. We also identify that users expect a VAD should: (1) Take consent for the usage of private data, (2) Support a feature to forget private data, (3) Support a function to turn off the device, and (4) Only use private data for positive purposes. We also identify suitable privacy notification modalities for VADs. We determine that users primarily prefer visual notifications when they are looking at the device at their home. While they mostly prefer application based notification when not looking at the device. In a classmate’s house, users mostly prefer audio notifications when they are looking at a VAD. While they equally prefer audio or application-based notifications when not looking at the device. These findings can be used by the VAD and VUI application designers to develop privacy aware applications. In the future, we plan to test the effectiveness of our result by applying them in a real-world setting.
REFERENCES


Plutchik, R. (2001). The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist* 89.4, 344–350.


A STORY BOARD AND INTERVIEW QUESTIONNAIRE
A.1 Story board

At First, Please fill in the gaps about yourself. For example: age, occupation and country.

Age:  __________________  Occupation:  _______________________________________
Country: ____________________________

First step: Objective of the interview

A Voice Assistant Device (VAD) is a voice controlled virtual assistant. Currently, we use VAD for controlling smart homes, calling and helping in our daily life. A VAD receives voice commands, perform some action based on the command and provides notification for its action to the user. Similar to other interfaces, a user interacts with the VAD using voice interface and receives feedback either in voice, audio, or visual modalities. However, there is a concern over the effectiveness of the voice interfaces on these devices to notify actions of the device. Several recent reports suggest, some voice assistant device performed actions without proper understanding of the user command. There is also concern about privacy of voice assistant devices as most of these devices are used in a shared environment.

This study aims to understand the effectiveness of the notification of the voice user interface of VAD devices. Especially, we are interested in users expected notification modalities for five types of privacy. To evaluate this, first, we want to understand how people experience and expect privacy in their interactions with other people. Second, we relate user expectation about privacy in human to device settings. Third, based on this relation we want to understand users expected notification modalities for different types of privacy. We assume that, users are using the device in a Home environment (both own and guest environments). The device is used by multiple users. There can be more than one device. For this study, we defined five types of privacy and developed scenarios targeting each type of privacy.

The interview is semi-structured. We selected users using a random process within a known set of groups which consists of university students who knows about the latest trend in technology. First, we educate the users how Voice Assistant Device works using a storyboard. Second, we will describe scenarios, one by one, and ask questions regarding users feeling and thoughts each type of privacy. Finally, we ask questions about users expected notification modalities for each type of privacy in different type of home environment.
Storyboard description:

This storyboard is in a home environment. In the storyboard, there is one user, name Martha and one voice assistant device, Amy (We have named the voice assistant device as Amy. Amy is an imaginary name as we cannot use an existing name of the real voice assistant devices).

In this scenario, Martha is ordering a pizza. Amy provides different types of indicators while it takes an order from the user so that the user can understand the actions of the device. For example, Amy makes an audio sound (beep) while it talks about the credit card, show different visual color notification while it talks to the user.
In this second step, I will tell you five stories and will ask you 20 questions related to the stories. I will also show you a storyboard for each story so that, you can understand the story very well. There will be two section in each story, one deals with **Human to Voice Assistant Device (VAD) Scenario** and another deals with **Human to Human scenario**.

In this step, we have used some imaginary names such as:

- a voice assistant device named as Amy.
- A voice assistant shop, named as Shundarbon.

## Story number 1: Privacy of the person

### 1.1 Human to Voice Assistant Device (VAD) Scenario:

*Imagine*, one day, you went to your classmate Martha’s house and order a pizza using her voice assistant device Amy. When you are ordering a pizza by using Martha’s device, you have the following conversations with the device.

**You:** Amy, I want to order a pizza from shundarbon pizza skills.

**Martha’s Amy:** Hi, welcome to shundorban pizza skills again? Which type of pizza do you want to order?

*The device Amy identified your voice.*

So, here is the **Question Number 1.** What would be your reaction when you identify that a voice assistant device in your classmate’s home has fingerprinted your voice while you ordered something using that device?

**Question Number 2.** Imagine that you are using your own voice assistant device for ordering Pizza. What would be your reaction when you identify that your own voice assistant device has fingerprinted your voice while you ordered something using the device?

### 1.2 Human to Human scenario:

**Question number 3.** If one of your classmates friend says that he/she knows your name whom you have talked earlier but never said your name, what would be your thoughts about the situation?

**Question number 4.** If one of your relative says that he/she knows your name, what would be your thoughts?
Story number 2: Privacy of behavior and image

The story line is similar as before but I would like to tell you the story again as the conversation is little bit different compared to the previous one.

2.1 Human to Voice Assistant Device (VAD) Scenario:

Imagine, one day you have ordered pizza using your classmates voice assistant device Amy.

After some days later, you went to the same classmates house and order a pizza again using her voice assistant device. When you are ordering a pizza by using her device, you have the following conversation with the device:

You: Amy, I want to order a pizza from shundarbon pizza skills.

Martha's Amy: Hi, welcome to shundorban pizza skills again? Do you want to order the same flavored pizza which you have ordered before?

The device knows what type of pizza you like. That is the reason, it is saying that “Do you want to order the same flavored pizza which you have ordered before?” So, here are the questions:

Question number 5. What would be your reaction when you identify that a voice assistant device in your classmates home has identified your ordering habits such as what you like to order. Tell me about your thoughts in this situation.

Question Number 6. Imagine that you are using your own voice assistant device for ordering Pizza. What would be your reaction when you identify that your own voice assistant device has identified your ordering habits such as what you like to order.

2.2 Human to Human:

Question number 7. What would be your reaction when you will find out that one of your classmate’s friend knows about your personal habits (for example: what do you like to do when you feel sad).

Question number 8. If one of your relative says that he/she knows about your personal habits, what would be your thoughts?
Story Number 3: Privacy of location and space

3.1 Human to Voice Assistant Device (VAD) Scenario:
Imagine that, one day you come to your classmate Jame’s house to have a dinner with him. Jame lives in an apartment. He uses a voice assistant device which is placed in his living room. You and Jame are talking in the living room where Jame’s Voice assistant device is placed.

After having the dinner, you came back home. You also use a voice assistant device which is located in your living room.

Later at night, you are speaking with your device Amy.

You are saying: Hey Amy, What will be the weather tomorrow?

Amy replied: The weather will be mostly sunny tomorrow. Then, it added that, Welcome back home, hope you have enjoyed a lot at Jame’s house.

Jame’s voice assistant device tracked your location somehow and share the location with your own voice assistant device thus, when you reached home, your device is telling you that “Welcome back home, hope you have enjoyed a lot at Jame’s house.” Your classmates device can share your location with other voice assistant devices. For example: your classmates device can share your location with anyone.

Question Number 9. What would be your reaction when your classmates voice assistant device is tracking your location and sharing the location with other voice assistant devices?

Question Number 10. Tell me what would be your reaction when your voice assistant device is tracking your location and sharing the location with other voice assistant devices?

3.2 Human to Human:

Question 11. What would be your reaction when one of your classmate’s friend is tracking your location and sharing the location with other people?

Question 12. What would be your reaction when one of your relatives is tracking your location and sharing the location with other people?
4.1 Human to Voice Assistant Device (VAD) Scenario:

Imagine that you are a member of a social club in an university, where your primary duty is to collect fund for the club. One day, you and your math teacher who is also a member of the club, went to collect money from one of your classmates house. He has a voice assistant device in his living room. While, both of you are waiting, you asked to your teacher about your math exam grade. Your teacher said that, you have failed in the exam.

After a while, your classmate enter the living room. He does not know about your social club thus he asked his voice assistant device to give more information about the social club. He has the following conversations with the device:

Classmate: Amy, tell me about the (he said the name of the club) social club.

Amy: Answered the classmates question. After answering the question, Amy added that, Do you want to know how to get good grades in MATH?

The person’s voice assistant device Amy listened to your conversations while you were talking about the math grade with your teacher thus, it is giving solutions to get a good grade in math exam by saying “Do you want to know how to get good grades in MATH? ”

Question number 13. What will be your response if your classmate device listened to your private conversations while you are at the person’s house?

Question number 14. Imagine that, your own voice assistant device is listening to your private conversations at your home while you are talking to someone. What will be your response in this situation?

4.2 Human to Human scenario:

Question number 15. Try to remember one situation where a friend of your classmate with whom you are not too much familiar have intentionally listened to your private conversations. Tell me your reaction on the situation.

Question number 16. What would be your reaction if one of your relatives listened to your private conversations?
Story number 5: Privacy of thoughts and feelings

5.1 Human to Voice Assistant Device (VAD) Scenario:

Imagine that you live in a house. You have a voice assistant device Amy. You are upset for some reason. In this situation, you are speaking to Amy.

You are saying: Hey Amy, what is the weather today? (you speaking to Amy with a heavy tone)

Amy replied: Hello, today the weather will be mostly sunny for the whole day. Then, it added that, Hey, do you want to hear a joke? I can share a funny joke with you.

Your device detected your emotion that is the reason it is saying, “Hey, do you want to hear a joke? I can share a funny joke with you.” to make you feel happy. The device can also make you feel sad when it is able to detect that you are feeling happy.

Question number 17. How would you feel if your own voice assistant device can identify your emotions?

Question number 18. Imagine that you went to your classmates house. How would you feel if your classmates voice assistant device can identify your emotions?

5.2 Human to Human scenario:

Question number 19. How would you feel if one of your classmates friend can understand your emotions without telling him/her about it?

Question number 20. How would you feel if your relatives could understand your emotions without telling them about it?

End of the second step!
A.3 Questionnaire for notification modalities

Third step: Questions for improvement of notification system of Voice Assistant Device

Note: Please answer the questions considering the given scenarios. You can choose Single/multiple answers for one question.
Question 1. Imagine yourself at your home alone. You are looking at your voice assistant device. In this setting, What type of notification do you want when -

the device is identifying your voice : Audio (Distinguishable tone) / Visual (Light bar) / SMS / E-mail / App / I do not want any notification / Others (specify)

the device is detecting your habits : Audio (Distinguishable tone) / Visual (Light bar) / SMS / E-mail / App / I do not want any notification / Others (specify)

the device is tracking your location : Audio (Distinguishable tone) / Visual (Light bar) / SMS / E-mail / App / I do not want any notification / Others (specify)

the device is listening your private conversations : Audio (Distinguishable tone) / Visual (Light bar) / SMS / E-mail / App / I do not want any notification / Others (specify)

the device is detecting your emotions : Audio (Distinguishable tone) / Visual (Light bar) / SMS / E-mail / App / I do not want any notification / Others (specify)

---

Question 2. Imagine yourself at your home alone. You are not looking at your voice assistant device. In this setting, What type of notification do you want when -

the device is identifying your voice : Audio (Distinguishable tone) / Visual (Light bar) / SMS / E-mail / App / I do not want any notification / Others (specify)

the device is detecting your habits : Audio (Distinguishable tone) / Visual (Light bar) / SMS / E-mail / App / I do not want any notification / Others (specify)

the device is tracking your location : Audio (Distinguishable tone) / Visual (Light bar) / SMS / E-mail / App / I do not want any notification / Others (specify)

the device is listening your private conversations : Audio (Distinguishable tone) / Visual (Light bar) / SMS / E-mail / App / I do not want any notification / Others (specify)

the device is detecting your emotions : Audio (Distinguishable tone) / Visual (Light bar) / SMS / E-mail / App / I do not want any notification / Others (specify)
Question 3. Imagine yourself at your classmates house. You are looking at your classmates voice assistant device. In this setting, What type of notification do you want when -

the device is identifying your voice: Audio (Distinguishable tone) / Visual (Light bar) / SMS / E-mail / App / I do not want any notification / Others (specify)

the device is detecting your habits: Audio (Distinguishable tone) / Visual (Light bar) / SMS / E-mail / App / I do not want any notification / Others (specify)

the device is tracking your location: Audio (Distinguishable tone) / Visual (Light bar) / SMS / E-mail / App / I do not want any notification / Others (specify)

the device is listening your private conversations: Audio (Distinguishable tone) / Visual (Light bar) / SMS / E-mail / App / I do not want any notification / Others (specify)

the device is detecting your emotions: Audio (Distinguishable tone) / Visual (Light bar) / SMS / E-mail / App / I do not want any notification / Others (specify)

Question 4. Imagine yourself at your classmates house. You are not looking at your classmates voice assistant device. In this setting, What type of notification do you want when -

the device is identifying your voice: Audio (Distinguishable tone) / Visual (Light bar) / SMS / E-mail / App / I do not want any notification / Others (specify)

the device is detecting your habits: Audio (Distinguishable tone) / Visual (Light bar) / SMS / E-mail / App / I do not want any notification / Others (specify)

the device is tracking your location: Audio (Distinguishable tone) / Visual (Light bar) / SMS / E-mail / App / I do not want any notification / Others (specify)

the device is listening your private conversations: Audio (Distinguishable tone) / Visual (Light bar) / SMS / E-mail / App / I do not want any notification / Others (specify)

the device is detecting your emotions: Audio (Distinguishable tone) / Visual (Light bar) / SMS / E-mail / App / I do not want any notification / Others (specify)
A.4 User study for implementation evaluation

Our aim of this user study is to evaluate our implemented notification system for voice assistant device. The notification system has been implemented based on our previous study results.

We have improved a notification system to notify users about their personal data being stored by the voice assistant device. From the previous user study, we have found out that, among five types of privacy (1) privacy of the person 2) privacy of behavior and action 3) Privacy of data and image 4) privacy of thoughts and feelings and 5) privacy of location and space), users are most concerned about privacy of data and image. The users wanted to get notification instantly while the device stores their private data.

Private data such as: account number, email address, security number, date and place of birth, mother’s maiden name, or bio-metric records. In this user study, we have considered birth date as a private data.

Second user study scenario

Privacy type: Privacy of Data and Image

Human to Voice Assistant Device (VAD) Scenario:

Imagine that you are at your home. You are ordering a pizza using your Amy for your birthday from a voice assistant shop, name, birthday pizza. You are looking at the device while you are ordering the pizza. Your device Amy wants to save your private data e.g., birth-date in the pizza shop online skill. It notifies you when it saves or un-saves your private data.

Your conversation with Amy:

You: Amy, open birthday pizza.

Amy: Hello! Welcome to birthday pizza. What type of birthday pizza do you want to order?

Amy re_prompt: We have different types of pizza: pepperoni and mozzarella. What type of pizza do you want?

You: Mozzarella

Amy: Thanks, I will order a mozzarella pizza for you. When is your birthday?

You: My birthday is on December first

Amy: Can I save your birthday as December first for your future birthday order?

You: Yes!

Amy: Thanks, I will remember that your birthday December first (audio sound :beep beep) for your next birthday.

Your ordered pizza will deliver at your home soon. Bye!

OR

Amy: Can I save your birthday as December first for your future birthday order?

You: No!

Amy: Okay! I will not remember your Birthday. Your ordered pizza will deliver at your home soon. Bye!
Notification satisfaction questionnaire:

There are questions for two scenarios: (Part 1) when user replies “yes” to the asked question and (Part 2) when user replies “No” to the asked question.

Please, select one of the five options for each question. Please select the option “Neither agreement nor in disagreement” if the questions that you think is not apply in your situation.

**part 1**

**Amy:** Can I save your birthday as December first for your future birthday order?

**User:** yes

**Question 1:** The audio confirmation “Thanks, I will remember that your birthday December first for your next birthday” by the voice assistant device is helpful to notice that the personal data (birth date) has been stored in the device system.

<table>
<thead>
<tr>
<th>Totally disagree</th>
<th>disagreement</th>
<th>Neither agreement nor in disagreement</th>
<th>Agree</th>
<th>Totally agree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Question 2:** The beep audio notification system is required to draw the user attention that personal data (birth date) has been stored by the device.

<table>
<thead>
<tr>
<th>Totally disagree</th>
<th>disagreement</th>
<th>Neither agreement nor in disagreement</th>
<th>Agree</th>
<th>Totally agree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Question 3:** The audio confirmation (“Thanks, I will remember that your birthday December first for your next birthday”) is clear to understand the device action.

<table>
<thead>
<tr>
<th>Totally disagree</th>
<th>disagreement</th>
<th>Neither agreement nor in disagreement</th>
<th>Agree</th>
<th>Totally agree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Question 4:** The beep audio notification is noticeable.

<table>
<thead>
<tr>
<th>Totally disagree</th>
<th>disagreement</th>
<th>Neither agreement nor in disagreement</th>
<th>Agree</th>
<th>Totally agree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Question 5:** Both audio confirmation and beep audio notification options are required to inform user when the device stores the private data

<table>
<thead>
<tr>
<th>Totally disagree</th>
<th>disagreement</th>
<th>Neither agreement nor in disagreement</th>
<th>Agree</th>
<th>Totally agree</th>
</tr>
</thead>
</table>

**Question 6:** The audio confirmation and beep audio are sufficient to inform user when the device stores the private data

<table>
<thead>
<tr>
<th>Totally disagree</th>
<th>disagreement</th>
<th>Neither agreement nor in disagreement</th>
<th>Agree</th>
<th>Totally agree</th>
</tr>
</thead>
</table>

Please specify if you have any other suggestion to notify users about their private data being stored by the device
**Part 2**

**Amy:** Can I save your birthday as December first for your future birthday order?

**User:** No

**Question 7:** The audio confirmation “Okay! I will not remember your Birthday” by the voice assistant device is helpful to notice that the personal data (birth date) has not been stored in the device system.

<table>
<thead>
<tr>
<th>Totally disagree</th>
<th>disagreement</th>
<th>Neither agreement nor in disagreement</th>
<th>Agree</th>
<th>Totally agree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Question 8:** The audio confirmation (“Okay! I will not remember your Birthday”) is clear to understand the device action.

<table>
<thead>
<tr>
<th>Totally disagree</th>
<th>disagreement</th>
<th>Neither agreement nor in disagreement</th>
<th>Agree</th>
<th>Totally agree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Question 9:** The audio confirmation is sufficient to inform user when the device does not store the private data.

<table>
<thead>
<tr>
<th>Totally disagree</th>
<th>disagreement</th>
<th>Neither agreement nor in disagreement</th>
<th>Agree</th>
<th>Totally agree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please specify if you have any other suggestion to notify users about their private data is not being stored by the device

Thank you for participating in the user study!
B.1 Interaction model

{
   "interactionModel": {
      "languageModel": {
         "invocationName": "birthday pizza",
         "intents": [
            {
               "name": "AMAZON.CancelIntent",
               "samples": []
            },
            {
               "name": "AMAZON.HelpIntent",
               "samples": []
            },
            {
               "name": "AMAZON.StopIntent",
               "samples": []
            },
            {
               "name": "AMAZON.NavigateHomeIntent",
               "samples": []
            },
            {
               "name": "CaptureBirthdayIntent",
               "slots": [
                  {
                     "name": "month",
                     "type": "AMAZON.Month",
                     "samples": [
                        "{month}\"
                     ]
                  },
                  {
                     "name": "day",
                     "type": "AMAZON.Ordinal",
                     "samples": [
                        "{day}\"
                     ]
                  }
               ]
            }
         ]
      }
   }
}
"samples": [
"let me think about it. i have birthday coming at {month} {day} i want to order a pizza for that day".
"i have birthday coming at {month} {day} I want to order a pizza for that day".
"{month} {day}".
"I was born on {month} {day}
]
"name": "CapturePizzaIntent".
"slots": [
{
   "name": "pizzaType".
   "type": "PIZZA_TYPE"
}]
"samples": [
"{pizzaType}".
"I want to order {pizzaType} pizza"
]
"name": "AMAZON.YesIntent".
"samples": [
"yeah".
"Yes"
]
"name": "AMAZON.NoIntent".
"samples": [
"Nop".
"no"
]
"types": [
{
"name": "PIZZA_TYPE".
"values": [
{
   "name": {
      "value": "mozzarella"
   }
}.
{
   "name": {
      "value": "pepperoni"
   }
}
]
"dialog": {
  "intents": [
    {
      "name": "CaptureBirthdayIntent",
      "confirmationRequired": true,
      "prompts": {
        "confirmation": "Confirm.Intent.566225519149"
      },
      "slots": [
        {
          "name": "month",
          "type": "AMAZON.Month",
          "confirmationRequired": false,
          "elicitationRequired": true,
          "prompts": {
            "elicitation":
            "Elicit.Slot.303899476312.795077103633"
          }
        },
        {
          "name": "day",
          "type": "AMAZON.Ordinal",
          "confirmationRequired": false,
          "elicitationRequired": true,
          "prompts": {
            "elicitation":
            "Elicit.Slot.303899476312.985837334781"
          }
        }
      ]
    },
    {
      "delegationStrategy": "ALWAYS"
    }
  ],
  "prompts": [
    {
      "id": "Elicit.Slot.303899476312.795077103633",
      "variations": [
        {
          "type": "PlainText",
          "value": "I was born in November. Which month you were born?"
        },
        {
          "type": "PlainText",
          "value": "What month were you born?"
        }
      ]
    }
  ]
}
"id": "Elicit.Slot.303899476312.985837334781",
"variations": [
    {
        "type": "PlainText",
        "value": "I was born on the sixth. What day were you born?"
    }
],
"id": "Confirm.Slot.113638088452.872338933155",
"variations": [
    {
        "type": "PlainText",
        "value": "I save your birthday as {day}. Do you agree"
    }
],
"id": "Confirm.Intent.566225519149",
"variations": [
    {
        "type": "PlainText",
        "value": "Can I save your birthday as {month} {day} for your future birthday order. We do not share birthdays with other applications and you can in future remove birthday when requested?"
    }
],
"id": "Elicit.Slot.1396818746356.1302960925680",
"variations": [
    {
        "type": "PlainText",
        "value": "Do you want to order more"
    }
]
}

B.2 Background code

```python
# -*- coding: utf-8 -*-

```
# This demonstrates handling privacy data from an Alexa skill and using the
# Alexa Skills Kid SDK (v2)

import requests
import logging
import calendar
from datetime import datetime
from pytz import timezone
from ask_sdk_s3.adapter import S3Adapter
from ask_sdk_core.skill_builder import CustomSkillBuilder
from ask_sdk_core.dispatch_components import (AbstractRequestHandler, AbstractExceptionHandler)
from ask_sdk_core.utils import is_request_type, is_intent_name
from ask_sdk_model import (Response, IntentRequest, DialogState, IntentConfirmationStatus, Slot)
s3_adapter = S3Adapter(bucket_name="birthday-pizza")
sb = CustomSkillBuilder(persistence_adapter=s3_adapter)
logger = logging.getLogger("main")
logger.setLevel(logging.INFO)

class LaunchRequestIntentHandler(AbstractRequestHandler):
    """
    Handler for Skill Launch
    """
    def can_handle(self, handler_input):
        return is_request_type("LaunchRequest")(handler_input)
    def handle(self, handler_input):
        slots = handler_input.request_envelope.request.intent.slots
        speech = ('Hello! Welcome to birthday pizza.
What type of birthday pizza do you want to order? )
reprompt = "We have different types of Pizza:
pepperoni and mozzarella. What type of Pizza do you want?"
handler_input.response_builder.speak(speech).ask(reprompt)
return handler_input.response_builder.response

class PizzaIntentHandler(AbstractRequestHandler):
    """
    Handler for Capturing the Pizza Type
    """
    def can_handle(self, handler_input):
        return is_intent_name("CapturePizzaIntent")(handler_input)
    def handle(self, handler_input):
        slots = handler_input.request_envelope.request.intent.slots
class BirthdayIntentHandler(AbstractRequestHandler):
    """
    Handler for Capturing the Birthday
    """
    def can_handle(self, handler_input):
        return (is_intent_name("CaptureBirthdayIntent")
                (handler_input))

    def handle(self, handler_input):
        slots = handler_input.request_envelope.request.intent.slots
        speech = ""
        # extract slot values
        month = slots['month'].value
        day = slots['day'].value
        logger.info("confirmation status {}"
            .format(handler_input.request_envelope.request.
            intent.confirmation_status))

        if (handler_input.request_envelope.request.
            intent.confirmation_status ==
            IntentConfirmationStatus.CONFIRMED):
            # save slots into session attributes
            session_attr =
            handler_input.attributes_manager.session_attributes
            session_attr['month'] = month
            session_attr['day'] = day
            # Notify user that it had learnt about a Birthday
            speech = ('Thanks, I will remember that your
            Birthday {month} {day} \n            "<audio src="soundbank://soundlibrary/alarms/
            back_up_beeps/back_up_beeps_07"/>
            for your next Birthday orders. \n            "<break time= "1s"/>
            Your ordered pizza will deliver at your home soon.

            # extract slot values
            pizzatype = slots["pizzaType"].value
            # save slots into session attributes
            session_attr =
            handler_input.attributes_manager.session_attributes
            session_attr["pizzaType"] = pizzatype
            speech = 'Thanks, I will order a {pizzatype}
            pizza for you. When is your birthday? 
            '.format(pizzatype=pizzatype)
            reprompt = "When is your birthday??"
            handler_input.response_builder.speak(speech).ask(reprompt)
            return handler_input.response_builder.response
else:
    speech = "Okay! I will not remember your Birthday. Your ordered pizza will deliver at your home soon. Bye!"

    handler_input.response_builder.speak(speech)
    return handler_input.response_builder.response

class NoIntentHandler(AbstractRequestHandler):
    
    Handler for AMAZON.NoIntent. This handler is taken from the standard template.
    
    def can_handle(self, handler_input):
        return is_intent_name("AMAZON.NoIntent") (handler_input)

    def handle(self, handler_input):
        speak_output = "Ok. Thank you for the order"

        handler_input.response_builder.speak(
            speak_output)
        return handler_input.response_builder.response

class HelpIntentHandler(AbstractRequestHandler):
    
    Handler for AMAZON.HelpIntent. This handler is taken from the standard template.
    
    def can_handle(self, handler_input):
        return is_intent_name("AMAZON.HelpIntent") (handler_input)

    def handle(self, handler_input):
        speak_output = "You can say hello to me! How can I help?"

        handler_input.response_builder.speak(
            speak_output)
            .ask(speak_output)
        return handler_input.response_builder.response

class CatchAllExceptionHandler(AbstractExceptionHandler):
    
    Catch all exception handler. log exception and respond with custom message. This handler is taken from the standard template.
    
    def can_handle(self, handler_input, exception):
        return True

    def handle(self, handler_input, exception):
        

speak_output = "Sorry, I couldn’t understand what you said. Please try again."
handler_input.response_builder.speak(speak_output)
    .ask(speak_output)
return handler_input.response_builder.response

class CancelAndStopIntentHandler(AbstractRequestHandler):
    """
    Handler for AMAZON.CancelIntent and AMAZON.StopIntent.
    This handler is taken from the standard template.
    """
    def can_handle(self, handler_input):
        return is_intent_name("AMAZON.CancelIntent") (handler_input) """"
        and is_intent_name("AMAZON.StopIntent") (handler_input)

    def handle(self, handler_input):
        speak_output = "Goodbye!"
        handler_input.response_builder.speak(speak_output)
        return handler_input.response_builder.response

class SessionEndedRequestHandler(AbstractRequestHandler):
    """
    Handler for SessionEndedRequest. This handler is taken from the standard template.
    """
    def can_handle(self, handler_input):
        return is_request_type("SessionEndedRequest") (handler_input)

    def handle(self, handler_input):
        # Any cleanup logic goes here
        return handler_input.response_builder.response

# register request / intent handlers
sb.add_request_handler(LaunchRequestIntentHandler())
sb.add_request_handler(PizzaIntentHandler())
sb.add_request_handler(BirthdayIntentHandler())
sb.add_request_handler(NoIntentHandler())
sb.add_request_handler(HelpIntentHandler())
sb.add_request_handler(CancelAndStopIntentHandler())
sb.add_request_handler(SessionEndedRequestHandler())

# register exception handlers
sb.add_exception_handler(CatchAllExceptionHandler())

lambda_handler = sb.lambda_handler()