An Exploration into How Users Perceive and Trust Voice Assistant Technology

Liam Rawsthorne

Submitted in partial fulfilment of
the requirements of Edinburgh Napier University
for the Degree of
MSc Computing (User Experience)

School of Computing

August 2020

Milestones	Date of completion	Target deadline
Proposal	12/06/2020	Week 3
Initial report	08/07/2020	Week 7
Full draft of the dissertation	23/08/2020	2 weeks before final deadline

Learning outcome	The markers will assess	Pages [1]	Hours spent
Learning outcome 1 Conduct a literature search using an appropriate range of information sources and produce a critical review of the findings.	* Range of materials; list of references * The literature review/exposition/background information chapter	79 - 87 16 - 27	-
Learning outcome 2 Demonstrate professional competence by sound project management and (a) by applying appropriate theoretical and practical computing concepts and techniques to a non-trivial problem, or (b) by undertaking an approved project of equivalent standard.	* Evidence of project management (Gantt chart, diary, etc.) * Depending on the topic: chapters on design, implementation, methods, experiments, results, etc.	91 28 - 65	-
Learning outcome 3 Show a capacity for self-appraisal by analysing the strengths and weakness of the project outcomes with reference to the initial objectives, and to the work of others.	* Chapter on evaluation (assessing your outcomes against the project aims and objectives) * Discussion of your project's output compared to the work of others.	74 - 78 66 - 73	-
Learning outcome 4 Provide evidence of the meeting learning outcomes 1-3 in the form of a dissertation which complies with the requirements of the School of Computing both in style and content.	* Is the dissertation well-written (academic writing style, grammatical), spell-checked, free of typos, neatly formatted. * Does the dissertation contain all relevant chapters, appendices, title and contents pages, etc. * Style and content of the dissertation.		-

Learning outcome 5 Defend the work orally at a viva voce examination.	* Performance * Confirm authorship	1 hour

Have you previously uploaded your dissertation to Turnitin?	No
Has your supervisor seen a full draft of the dissertation before submission?	Yes
Has your supervisor said that you are ready to submit the dissertation?	Yes

Authorship Declaration

I, **Liam Rawsthorne**, confirm that this dissertation and the work presented in it are my own achievement.

Where I have consulted the published work of others this is always clearly attributed;

Where I have quoted from the work of others the source is always given. With the exception of such quotations this dissertation is entirely my own work;

I have acknowledged all main sources of help;

If my research follows on from previous work or is part of a larger collaborative research project I have made clear exactly what was done by others and what I have contributed myself;

I have read and understand the penalties associated with Academic Misconduct.

I also confirm that I have obtained **informed consent** from all people I have involved in the work in this dissertation following the School's ethical guidelines

Signed: **Liam Rawsthorne**

Date: 10/08/2020

Matriculation no: 40085161

General Data Protection Regulation Declaration

Under the General Data Protection Regulation (GDPR) (EU) 2016/679, the University cannot disclose your grade to an unauthorised person. However, other students benefit from studying dissertations that have their grades attached.

Please sign your name below *one* of the options below to state your preference.

The University may make this dissertation, with indicative grade, available to others.

Liam Rawsthorne

The University may make this dissertation available to others, but the grade may not be disclosed.

The University may not make this dissertation available to others.

Abstract

Voice assistants are a form of conversational artificial intelligence that is growing in popularity. From their introduction in smartphones and smart speakers, they have found their place in devices such as smartwatches, televisions and even transportation. As they grow in popularity, how humans feel towards the technology becomes important. Based on a review of literature in the domain of conversational artificial intelligence, a mixed-methods approach was taken to explore the trust and perception that frequent users and infrequent users have towards voice assistants when changing how the voice assistant sounds as well as the relationship between the different variables.

Discussions revealed themes and ideas that showed frequent users and infrequent users had similar thoughts, frequent users were more accepting of the concerns held by infrequent users and exploring this further suggests that there is no significant difference in how both groups perceive voice assistants. There is a clear difference, though, in how frequent users and infrequent users trust voice assistants. After using the personalised assistant for a week, we see this gap in trust grow further with frequent users seeing an increase and infrequent users showing a decrease. This is also seen in the users' perception of the assistants, where again frequent users saw a more positive perception while infrequent users' saw a more negative perception. The relationship between trust and perception proved to be a positive one suggesting that if a user perceives an assistant positively then they will trust it more. The relationship between trust speech-system usability is also positive, with accuracy and likeability showing the strongest correlations suggesting that if voice assistants provide accurate responses, and are pleasant to interact with then users will show higher trust.

List of Tables and Figures

Table of Contents

1. Introduction

Artificial intelligence (AI) was introduced as a concept in the first half of the 20th century. It was explored by scientists and mathematicians who argued that if a human could use information to complete tasks then machines would be able to do it too (Turing, 1950). Unfortunately, computers were not advanced enough, they could carry out commands but could not store them. Logic Theorist was then developed, and became the first programme that could execute and store commands.

Since then it has emerged in a number of fields ("Where is Artificial Intelligence Used Today?", 2020) such as healthcare (Davenport & Kalakota, 2019), security and surveillance (Triantafyllidou & Tefas, 2016), and self-driving vehicles (Rajkumar, 2015), but specifically, the field of conversational artificial intelligence (CAI). CAI is a form of AI that engages in conversation with humans with the aim of being as natural as possible. One form of CAI that has gained mainstream popularity recently is voice assistant technology.

1.1. Research Background

A Voice Assistant, also known as conversational agent (Clark, Munteanu, Wade, Cowan, Pantidi & Cooney, 2019) or conversational user interface (CUI) (Reeves, Porcheron, Fischer, Candello, McMillan, McGregor, Moore, Sikveland, Taylor, Velkovska & Zouinar, 2018) is a form of AI that uses voice recognition and natural language processing (NLP) in order to provide information or perform tasks for the user interacting with them ("How Voice Assistants Are Changing Our Lives | Smartsheet", 2020). They are predominantly integrated into devices within the home such as smart speakers, TVs, tablets and computers (Porcheron, Fischer, Reeves & Sharples, 2018) but also in devices that we would normally take out of the home such as smartphones, smartwatches and even in cars. They allow users to make requests such as setting alarms, asking for the weather and controlling music. This voice based, hands-free approach allows users to make these requests without interrupting a current task. A user may be focusing on writing a document, and without getting up or switching windows can have a query solved in seconds offering a higher level of productivity. They are also used in situations where a user's hands may be controlling something else such as when driving (Lahoual & Frejus, 2019) and their eyes must be

focused on the road, or in the kitchen (Porcheron et al., 2018) where hands may get messy and cannot be used on a device. There has been research into how people use their voice assistants and what they do with them, (Sciuto, Saini, Forlizzi & Hong 2018), (Porcheron et al., 2018), however there is little research into trust that users have for voice assistant technology (Edwards & Sanoubari, 2019).

1.1.1. Why Voice Assistants?

The growing popularity of voice assistants is a reason why the topic was selected as the focus of this dissertation. In 2019 it was estimated that, worldwide, 3.25 billion devices used a voice assistant and in 2023 this is predicted to rise to around 8 billion (Statista Research Department, 2019) showing the expected growth of the relatively new technology. It is also expected to grow beyond personal/domestic use with plans already in place to introduce voice assistants into workplaces such as offices (Finnegan, 2020) hospitals (Brooks & Heuser, 2020) as well as education (Plummer, n.d.). As voice assistants are implemented and used in other fields and industries, the number of people using them will continue to increase.

1.2. Aims and Objectives

1.2.1. Aims

The main aim of this research is to explore trust and perception that users show towards voice assistants.

One aim is to investigate the field of CAI and gain an understanding into how people use voice assistants as well as the factors that could affect their experience of using the technology, and this information will be collected through the exploration of current research in the form of a literature review.

Another aim is to discover the differences between frequent users and infrequent users by understanding their perceptions of voice assistant technology. This will be achieved through focus groups/interviews with frequent users and infrequent users, the data will then be analysed using thematic analysis. A questionnaire will also be completed to quantify this perception and the same questionnaire will be completed later on to show changes. Descriptive statistics and statistical analysis will help to analyse the data.

Another aim will be to understand trust in frequent users and infrequent users and how this can develop. This will be achieved by participants taking part in a week-long study where they will change the voice of their voice assistant. A questionnaire will be taken at the start to measure their initial trust with the default voice settings and the same questionnaire will be taken after a week to measure how their initial trust has changed after changing the voice settings. Descriptive statistics and statistical analysis will, again, help to analyse the data.

Finally, relationships with trust will also be explored. The previously measured trust will be compared with the previously measured perception. A questionnaire will be taken to measure speech-system usability which will then be compared with user trust. Again, descriptive statistics and statistical analysis will assist in analysing the data.

1.2.2. Objectives

- Carry out a literature review on voice assistants and the conversational user experience (CUX) that surrounds them.
- Gather participants to collect qualitative and quantitative data regarding their perceptions of trust and social presence/humanness
- Understand trust and changes in it between frequent users and infrequent users
- Evaluate the speech-system usability of voice assistants

1.3. Dissertation Structure

This dissertation is separated into the following five chapters:

- **Literature Review -** This chapter will explore the current research in regards to the factors such as trust and social presence that affect the usage of voice assistants
- **Methodology -** This chapter will explain different types of research as well as the type selected for this study. It will provide information about participants, the methods and tools used, and how these methods and tools were implemented into the research
- **Analysis -** This chapter will show analysis of the collected data

- **Discussion -** This chapter will discuss the analysed data as well as reviewing it in regards to the literature review
- **Conclusion -** This chapter will summarise the project as well as considering any future work that could possibly be conducted

2. Literature Review

This chapter will explore current research in the field of CAI. It will explore the conversational user experience (CUX) and explore the topic of trust and how that may prevent people from integrating voice assistants into their life. Social presence will also be examined and will look at how non-human products can be enhanced to come across as more humanlike.

2.1. Conversational User Experience

User experience (UX) is a term to describe the experience of using things such as a product, website, app or service. It “encompasses all aspects of the end-user’s interaction with the company, its services, and its products” (Norman & Nielsen, n.d.) and when applied to conversational user interfaces (such as voice assistants, text-based systems or social robots) it can be described as conversational user experience. When creating a CUI, the CUX (Sayago et al., 2019) must be considered. The CUX is the UX of any natural language based technology that replicates a human conversation and good CUX should make the flow between two people feel natural, even if one is a virtual assistant or a chatbot (Morphy, 2018).

Due to the growth in popularity of CAIs more people are integrating them into everyday life in the same way they would humans (Sundar, Jung, Waddell & Kim, 2017). When designing a CUX, understanding how human conversation is structured is necessary, the CAI should be designed to make the experience as realistic as talking to a human and to be as natural as possible (Moore & Arar, 2019) however some may argue that going beyond this could be devastating (Aylett, Cowan and Clark, 2019).

The CUX is an important aspect of the human-computer interaction (HCI) between users and their voice assistants.

2.2. Current Research in Conversational User Experience of Voice Assistants

Current research on voice assistant appears to focus on the quality of the voice assistants and how users use their voice assistants.

In one study, four different voice assistants (Google Assistant, Alexa, Siri and Cortana) were evaluated in terms of how accurate the responses to general questions were and whether the answer could be considered a “good answer” (Berdasco, López, Diaz, Quesada & Guerrero, 2019). They found that Google Assistant and Alexa were much better than Cortana and Siri (with Siri performing the worst) in providing a correct answer with the same result showing for quality of the answer. Users were much more impressed with the performance of Google Assistant and Alexa (describing both as excellent in regards to “quality” and “correctness”) compared with the other two (Cortana was described as “above average” for both whereas Siri was described as ”average” in the two categories).

Another study explored the quality of answers provided by voice assistants however this time focused on health questions (Alagha & Helbing, 2019). Similar to the previous study, they found Google Assistant was a high performer, however, there were contrasting results for Siri and Alexa. Siri scored the highest with a score of 5.16 out of 6 whereas Alexa performed the worst scoring 0.98 out of 6. These scores were based on how well it understood the request as well as the quality of its answer. Google Assistant and Siri both understood the questions and provided an answer 94% of the time whereas Alexa only did this 25% of the time. Google Assistant and Siri both cited from a reliable source, CDC.gov (Centers for Disease Control and Prevention), the most whereas Alexa mainly cited from Wikipedia, an unreliable source. They state that this could be due to the search engine used (Google Assistant and Siri use Google, and Alexa uses Microsoft Bing) and that these search engines have had medical information partnerships in the past. Results differ from the previous study (Berdasco et al., 2019) and it appears that the system’s quality could depend on the type of information that is being requested (Alexa was poor in regards to health

questions, however in regards to general trivia it scored very high) (Alagha & Helbing, 2019).

Another study focusing on pragmatic quality (task-orientated) and hedonic quality (appealing and fun) (Kocabalil et al., 2018) is a comparison between Google Assistant and Lyra (Orehovacki, Babic & Etinger, 2017). They find that Lyra significantly performs better than Google Assistant in both pragmatic and hedonic quality. In the previous study (Berdasco et al., 2019), they found that Google Assistant was the best in providing quality responses however here it is discovered that Google Assistant was significantly worse than a less known voice assistant. Research mainly focuses on the same voice assistants due to their size and background, however this shows that smaller assistants should possibly be considered also.

Some studies however have focused on the user rather than the system itself such as when exploring user satisfaction through understanding it (Kiseleva, Williams, Jiang, Hassan Awadallah, Crook, Zitouni & Anastaskos, 2016) or in their own follow up study, by trying to predict it (Kiseleva, Williams, Jiang, Hassan Awadallah, Crook, Zitouni & Anastaskos, 2016). Another example is of user engagement, where one study showed success predicting whether users will continue to use their voice assistants in the future (Sano, Kaji & Sassano, 2016) as well as another study proposing how to elicit a stonger sense of user engagement (Shi, Yan, Ma, Lou & Cao, 2018).

Other studies have explored how voice assistants are used in the home (Porcheron et al., 2018) as well as when there are multiple occupants of the same household interacting with it. One study claims that voice assistants caused interruption and were an interference requiring repeated supervision and verification (Lahoual & Frejus, 2019) whereas another study showed mainly positive experiences from households (Sciuto et al., 2018).

While the household studies included experiences of the individuals in the home, research has also explored specific and contrasting use groups too of these individuals. Studies have tried to understand how children use voice assistants (Lovato & Piper, 2015), (Druga, Williams, Breazeal & Resnick, 2017) as well as how the elderly function with them (Sayago, Neves & Cowan, 2019). Other opposition user

groups that have also been looked at are people who choose to use (Luger & Sellen, 2016) voice assistants and those who choose not to (Cowan et al., 2017).

While there has been several studies exploring the qualities of voice assistants and the interactions with them there has been little research into the trust that a user may have for the assistant. Some have touched on it and briefly mentioned it, however there has not been a large focus into what trust means and how it can be defined in regards to voice assistant technology (Edwards & Sanoubari, 2019).

2.3. Trust

Trust is an important issue that affects whether or not someone uses a piece of technology. It is the belief that one has for another party to do what is expected of them and defined as 'the willingness of a party to be vulnerable to the actions of another party' (Mayer, Davis & Schoorman, 1995). Through their examination of prior research into what affects trust, and their model proposed that the three main factors that mainly affect trust are ability, benevolence and integrity (Mayer et al., 1995). Ability refers to the skills and knowledge of the trustee; benevolence refers to the belief that the trustee is wanting to help the trustor; and integrity refers to whether the trustee meets the trustor's ethics to a respectable level. (Mayer et al., 1995). This is one model of interpersonal trust, however trust has been defined in areas within HCI such as automation (Lee & See, 2004) or ecommerce (Egger, 2000) (Edwards & Sanoubari, 2019).

One area of trust that has been researched in HCI is online trust. It has mainly existed with websites; however, similarities can be noted with voice assistants due to their common goal of supplying the user with information/entertainment. One review into online trust states that there are three main factors, with their own variables that affect trust: customer based, system based and organisation based (Beldad, de Jong & Steehouder, 2010)

- Customer based refers to the user's own tendency to trust as well as their experience of using the technology
- System based refers to the quality of information provided, how easy the system is to use, personalisation, privacy and security and social presence
- Organisation based refers to their reputation and familiarity (Beldad et al., 2010)

Despite research of trust in other fields within HCI, it is yet to be defined in the domain of CAIs and is an area that is under researched (Edwards & Sanoubari, 2019). This is surprising considering there has been several studies revealing that people do show concern in trusting CUIs (Cowan, Pantidi, Coyle, Morrissey, Clarke, Al-Shehri, Earley & Bandeira, 2017), (Clark et al., 2019), (Microsoft, 2019). There has been some

research however which has attempted to define trust in CAIs in regards to privacy (Clark et al., 2019) and the quality of the system (Luger & Sellen, 2016).

2.3.1. Privacy

To use a CAI a user must give up their personal data in the form of text or audio and be able to trust that their data is being handled correctly, however issues surrounding privacy have been brought up in regards to what happens after the interaction has taken place. When a user sends a request (e.g. “Alexa, add toilet roll to my shopping basket”) to their voice assistant the audio, in this instance, has been sent to Amazon. What happens afterwards with that recording? This has led researchers to identify the main aspects of these issues as data ownership (Cowan et al., 2017), data storage (Malkin, Deatrick, Tong, Wijesekera, Egelman & Wagner, 2019), (Cowan et al., 2017) and possible abuse of privacy (Lau, Zimmerman & Schaub, 2018).

In a study by the University of California, users were asked multiple questions regarding the storage of their data. They found that half of their participants did not know that their recordings were being stored permanently and that these files could be accessed, and only a quarter knew about Google and Amazon retaining the audio (Malkin et al., 2019). Upon discovering this many participants were upset in the permanency of the data but showed more discomfort in knowing that other users’ data was being stored when compared to the storage of their own. There is also concern around accidental recordings. Voice assistants are supposed to only respond to their wake-word (OK Google, Alexa, Hey Siri) however they occasionally pick up on audio that does not contain the prompt and 10% of audio that was examined was accidentally recorded (Malkin et al., 2019). Participants who did report privacy concerns said it was often because of these accidental recordings. Despite referring to these as accidental, some may start to question just how accidental they are: with 41% of users believing that their voice assistant is actively listening to or recording them (Microsoft, 2019).

This could prevent families or households with multiple occupants in adopting voice recognition devices. They may want to convert their homes into smart homes (Strengers, Kennedy, Arcari, Nicholls & Gregg, 2019) as the technology advances and

while the primary user may be fine with their voice being recorded they may have problems with the voice assistant possibly picking up on the voice of a child, a guest or anyone that should not be recorded. They may also feel uneasy with their voice being accidentally recorded in case sensitive information out is picked up on.

The uneasiness around the ownership and storage of their data has led users to show concern in what companies could then do with it and the possible abuse that could take place (Lau et al., 2018). Issues revolve around data breaches and the monetisation of data. Users may think their data will be stored securely and kept between themselves and the company, however, news stories have highlighted that this is not always the case. In one instance users' voices were recorded and sent to a random person in their contact list (Kim, 2020). Another example shows Google employees actively leaking the data themselves, with 15% of the recordings being audio that should not have been recorded as it did not follow the wake word (Siegal, 2019).

2.3.2. System Quality

Another area where a user must be able to trust the voice assistant is when relying on the system to function at an acceptable standard. If one is using a voice assistant to complete an objective such as selecting a playlist or finding a recipe, they are likely doing it for convenience (Smith, 2018). A voice assistant aims to make the life of the user better by improving productivity, but if it fails to complete tasks then it is ultimately useless. One user pointed out that their reluctance to use a voice assistant stemmed from not being able to trust that it could carry out a simple task properly. They stated that even if they ask their voice assistant to set an alarm then they would check their phone to make sure the alarm had been set, and if they're opening the app on the device anyway then what's the point of asking the voice assistant to do it (Porcheron et al., 2018)?

The voice assistant also must be able to provide correct information too. The user, as stated before, will likely be using the voice assistant for convenience, and like searching via any other method, they will be looking for the information provided to be correct and up to date. As seen in a previous study (Alagha & Helbing, 2019), the

quality of the voice assistant was partly based on the quality of information where the two voice assistants that mainly cited from a government organisation scored higher than the voice assistant that mainly cited from Wikipedia. It has been shown that users tend to trust a system more if the supplied content accurate and of a high quality (Sillence, Briggs, Harris & Fishwick, 2007).

While one study (Berdasco et al., 2019) did not focus on trust, the assistants' high quality answers do suggest that after using the voice assistants and understanding their responses, that high levels of trust could exist amongst the participants in the voice assistants that performed best. (Berdasco et al., 2019). Similarly, this could also be true in the similar study (Alagha & Helbing, 2019) into the responses of health questions.

2.3.3. Summary

In regards to the CUX, it could be argued that trust is the most important factor to affect it. It is the starting point of a user's journey with a voice assistant and if they cannot get past the doubt they have for the technology then there is little chance that they will use it. If the user can trust it, in terms of privacy, then there is the next step of trusting that the system can carry out its functionality. If they cannot trust the voice assistant to carry out tasks to a sufficient level, provide high quality information and be usable, it will be deemed useless and the user may completely give up using it.

2.4. Social Presence and Humanlike Characteristics

Social presence is the feeling of existing in a virtual environment as well as "the experience of sensing a social entity when interacting with the system" (Heerink, Krose, Evers & Wielinga, 2009). It is a concept that has been applied to area of human-robot interaction (HRI) such as in social robotics (Heerink et al., 2009) and while voice assistants currently show no physical resemblance to a social being in the same way robots do, they can be considered social entities. Voice assistants somewhat possess humanlike qualities, and several studies have explored the modification of these as having an effect on social presence such as embodiment, as well as exploring voice assistants with robotics.

As stated previously, an aim of CAI is to make the CUX feel natural as an interaction between humans (Moore & Arar, 2019) and in one study users were asked their opinions on the human-like nature of Siri. Some attached human attributes to the voice assistant such as discussing its personality and human-like voice and felt that this made it friendlier. The personality of the voice assistant is enhanced further with the addition of cultural influences. The user can change the accent/voice of the voice assistant, which in turn affects how it responds to the user. One user pointed out the British setting sounded posh, and that the Japanese setting was more serious, allowing the user to customise their own experience (Moore & Arar, 2019). This customisation would create a different experience for users, possibly a more positive one. People have shown to be more comfortable and trusting when with people from a similar cultural background and allowing to change (i.e. an Australian accent to a Spanish accent) could positively affect their CUX (McGill University, 2018).

The previous research mentions the alteration of the voice's accent and how this could be affected by cultural background, however one study explored the effects of other personality changes: by changing the voice assistant's language (low-status/high-status) and its gender (Habler, Schwind & Henze, 2019). They found that low-status language received more positive feedback with participants than high-status language and while users showed preference for a male voice, they noted that the gender that achieved highest performance scores was female, combined with low-status

language. This study suggests that personalising the voice assistant's personality does indeed change how users view the system.

Research has also examined the addition of embodiment to increase the human-like nature of voice assistants. While the current voice assistants focus on audible commands and responses, an important part of social interaction they lack is the ability to provide and understand social cues. This was explored in one study through the use of Augmented Reality (AR) to provide a visual body to the voice assistant (Kim, Boelling, Haesler, Bailenson, Bruder & Welch, 2018). The user makes a request and through an AR headset, the embodied assistant provides a social cue (such as leaving the room or putting on headphones and closing its eyes when privacy is asked for) to show the request has been acknowledged. With this embodiment and the added social behaviours, they found that participants recorded an improved sense of social presence (Kim et al., 2018). This is further backed up in another study where they found that a humanlike visualisation was preferred to an invisible agent and a wireframe of a human due to the social cues (eye contact and gaze) that could be provided (Reinhardt, Hillen & Wolf, 2020).

Another study researched using AR to provide a visual representation of the voice assistant's voice. They provided a selection of agents (humanlike, voice and non-human) to find out which one users enjoyed using and they found that users showed more of a preference towards the human visualisation (Wang et al., 2019). As voice assistants advance, this is something that could be considered in future versions.

While the previous study focused on embodiment through an AR headset, one study has compared users' perceptions of a robot and a voice assistant in carrying out the same activity (Pollmann, Ruff, Vetter & Zimmermann, 2020). Pragmatic quality and hedonic quality is compared between a voice assistant, a robot (named Pepper) without animation and the same robot with animation and the study shows that pragmatic quality is highest and hedonic quality is lowest with the voice assistant, whereas pragmatic quality is lowest and hedonic quality is highest with the animated robot. They also discovered that the robot without animation scored in the middle for both pragmatic and hedonic quality, with pragmatic only scoring 0.5 less (on a 1-7 scale) than the voice assistant. While the researchers were more concerned with fun

and entertainment, it does show that there was not much of a difference between a voice assistant and an unanimated robot in terms of a task-orientated nature (pragmatic) and that the fun and appeal (hedonic) towards it actually increased by changing how humanoid the system was (Pollmann et al., 2020). This suggests that the experience of social presence could be positively altered by adjusting the humanlike characteristics of voice assistants.

Pepper was used again in a study measuring trust between humans and robots. Participants took part in a pairs-matching game with Pepper (a humanoid robot that can communicate visually and verbally) and another robot called Husky (more machinelike with visual communication out without verbal communication). They found Pepper to be significantly more trustworthy than Husky and conclude that anthropomorphising a non-humanlike product does increase human trust; however, they do acknowledge that the human interacting with the device also has an effect on the outcome (Yaseen & Lohan, 2018).

Despite the idea of a more human-like machine appearing to show benefits, research has shown that this could be dangerous. We may think that if a voice assistant is trying to replicate human conversation, then it needs to act like a human, however this may not always be a positive thing with some research claiming that a non-human product gets creepier the more humanlike it becomes. It's argued that they may even sound more dishonest. (Aylett, Cowan & Clark, 2019). Personalising voice assistants may be acceptable to an extent, however taking too many steps to this could lead to growing negativity towards the systems. Other research has also shown that people did not want to talk to their voice assistant as if it was human (Clark et al., 2019) however one study shows that some users do in fact wish to converse with their assistant as if they were a human but acknowledge that their participant size is too small to conclude anything (Völkel, Kempf & Hussmann, 2020).

Through interface advancements, we are no longer only using graphical user interfaces (GUIs) to carry out everyday tasks, but we are now increasing our adoption of voice user interfaces (VUIs) (Myers, Grethlein, Furqan, Ontañón & Zhu, 2019). When interacting with an assistant device such as Google Home or Apple Homepod, we are inputting audio and receiving audio back. However, some devices, such as

Google Nest Hub or the Echo Show, are combining the VUI and the GUI together for the user to hear and see a response. This could provide a platform to give the voice an on-screen persona. As research (Kim et al., 2018) shows, embodiment and social behaviours increased social presence therefore displaying a visual body may be an improvement to the voice assistants that combine VUIs and GUIs.

2.4.1. Summary

To summarise, there are ways to enhance a user's experience by changing the human-like nature of a voice assistant, and as the studies in section show that on several occasions this change has been a positive one, with users' social presence improving as well as users gaining trust. However, it may only be beneficial to a certain point considering concepts like uncanny valley unsettling users. Further research will help solve this.

3. Methodology

This chapter will introduce the methodology by providing an explanation for the selected approach as well as describing other approaches that could have been considered to explore the following aims:

- This study aims to understand the differences between frequent users and infrequent users in their perception of voice assistants and how this perception changes.
- It also aims to understand the differences between frequent users and infrequent users in how they trust voice assistants, as well as how this trust changes.
- These perceptions and trust will also be compared with each other to understand if there is a relationship between perception and trust. Finally, the study also aims to explore if there is also a relationship between trust and speech-system usability.

This chapter will also explain the methods and tools used to achieve these aims, as well as describing how these methods and tools were implemented into the study. Information on the participants will also be provided.

3.1. Types of Research

When designing research, the researcher must recognise what they are studying, why they are doing so, the type of data they are attempting to collect, and whether this data is qualitative or quantitative. This will allow them to select the most appropriate approach.

Initially it was decided that a qualitative approach would be most appropriate as this research is concerned with understanding the experiences users have when interacting with a VA. There are five main types of qualitative research: ethnography, narrative, phenomenological, grounded theory, and case study. (Sauro, 2015)

- **Ethnography:** focuses on context and culture
- **Narrative:** understands individual experiences
- **Phenomenological:** focuses on describing an event or phenomenon

- **Grounded Theory:** provides the reasons behind an event
- **Case Study:** provides an in-depth focus on one test subject (Sauro, 2015)

For these reasons, a phenomenological approach was considered. This research is concerned with the experiences people have with VAs and understanding why they either use one or do not use one, thus using a framework that focuses on describing a phenomenon would be most appropriate. To optimise the research, it was decided that quantitative data should also be used alongside qualitative data. This resulted in the decision to use a mixed methods framework instead. Research is usually only referred to as quantative or qualitative, however the two can be combined to produce a mixed methods approach. Within this there are four main design types (Shorten & Smith, 2017).

- **Explanatory sequential:** begins with the collection and analysis of quantitative data followed by the collection and analysis of qualitative data to assist in explaining the initial data.
- **Exploratory sequential:** begins with the collection and analysis of qualitative data followed by the collection of quantitative data to test the initial findings
- **Parallel:** both qualitative and quantitative data are collected and analysed at the same time
- **Nested:** begins with either qualitative or quantitative data and the other is used to answer another question alongside the main research question (Shorten & Smith, 2017).

3.2. Approach

As stated previously, a phenomenological approach was considered, however, the decision to include quantitative research in the study led to an exploratory sequential approach being selected as its purpose is to explore a phenomenon using qualitative methods with the addition of quantitative results (Creswell, 2009).

As this piece of research has looked at how trust and perception changed over time, it can be described as a longitudinal study as opposed to a cross-sectional study. A cross-sectional study compares one group at one single point time with another group

at a different single point in time or how many different groups and times they would like; however, a longitudinal study focuses on one group over many different points in a certain amount of time. This allows researchers to explore how the group develops and any changes over a longer period ("Cross-sectional vs. longitudinal studies", 2015).

3.3. Participants

To begin the research, twelve participants (as seen in figure 1) between the ages of 23 and 58 (mean = 28.67) (eight male participants and four female participants) were recruited and divided into two groups of six “frequent users” (FUs) and six “infrequent users” (IUs). Rather than recruit participants and then divide them into the two groups, it was planned that each group would have a similar number of participants, therefore when recruiting participants, it was important to know whether or not they were a frequent user or an infrequent user. People were simply asked “would you call yourself a user of voice assistants (Alexa, Siri, Google Assistant etc.)?” and if they had access to a voice assistant, they were asked if they would take part in the study. Sample size was determined based on other studies that have looked at trust in voice assistants with similar numbers (Luger & Sellen, 2016), (Cowan et al., 2017). Having more participants was desired, however, due to restrictions surrounding coronavirus, this was not as easy to achieve.

Frequent users were considered “frequent” if they considered it a part of their everyday lives and infrequent users were considered “infrequent” if they had tried using it in the past but would not use it often, there was no in-between. Age and gender did not matter for the study, as they had no impact on how experienced they were, but they were collected for demographic purposes. The type of voice assistant participants used was also not required to be a specific type however the majority were owners of Google Assistants (Google Assistant = 7, Alexa = 4, Siri = 1). The two groups were split this way to allow for the findings from the two sets to be compared with each other.

All participants took part in every stage of the research.

Table 1

Participant Information

Participant	Gender	Age	Voice Assistant
P1	Male	29	Alexa
P2	Male	26	Google Assistant
P3	Male	27	Google Assistant
P4	Male	27	Google Assistant
P5	Male	29	Siri
P6	Male	26	Google Assistant
P7	Female	25	Google Assistant
P8	Male	26	Google Assistant
P9	Female	28	Alexa
P10	Female	25	Google Assistant
P11	Female	23	Alexa
P12	Male	58	Alexa

Note: This table shows information about participants of the study

3.4. Research Methods

To investigate the research questions, the appropriate methods and tools are necessary to gather accurate and useful data. To gather quantitative data, three questionnaires were utilised: Godspeed Questionnaire (Bartneck, Kulic, Croft & Zoghbi, 2008), The Checklist for Trust between People and Automation (Jian, Bisantz & Drury, 2000) and the Subjective Assessment for Speech System Interfaces (SASSI) (Hone & Graham, 2000). To gather qualitative data: focus groups and interviews were conducted to ask about experiences beforehand as well as afterwards. To analyse the data: descriptive statistics, statistical analysis and thematic analysis were completed.

All quantitative tools chosen have been published and are reliable to use for data collection. They can be trusted and is why they have been selected for use in this study. All qualitative tools were chosen as they would provide useful data, while remaining structured enough to put participants at ease to gather further responses. All analytical tools have been selected as the most appropriate tools needed for the desired results.

3.4.1. Godspeed Questionnaire

The Godspeed Questionnaire (Bartneck et al., 2008) is a tool that was developed to help in the progress of HRI While voice assistants would not be considered robots, they do share similarities which allow for the questionnaire to provide insightful results. It has twenty-four statements measured by Likert scales ranging from 1 to 5 separated into five sections: anthropomorphism, animacy, likeability, perceived intelligence and perceived safety.

- Anthropomorphism refers to the attributes normally seen in humans and applying them to something that is non-human product like a robot or a computer.
- Animacy is concerned with the lifelikeness of a non-human product.
- Likeability refers to a person's impression of the product
- Perceived intelligence is concerned with how smart a person believes a product to be

- Perceived safety refers to how comfortable a person feels when using the product as well as whether or not they feel in danger when using it (Bartneck et al., 2008)

One change has been made to the questionnaire due to the feedback provided directly to the original researcher via their own website. Several researchers had brought up the point that the final statement of “quiescent/surprised” was negatively correlated with the other two in its section ((anxious - relaxed) and (agitated - calm)). Therefore, it has been switched around to become “surprised/quiescent” to align correctly with them. This means that 5 is the optimal score when analysing the results.

3.4.1.1. Sample

A sample is shown below, the full questionnaire is available in Appendix A.

Godspeed II: Animacy

Please rate your impression of the robot on these scales:

Dead	1	2	3	4	5	Alive
Stagnant	1	2	3	4	5	Lively
Mechanical	1	2	3	4	5	Organic
Artificial	1	2	3	4	5	Lifelike
Inert	1	2	3	4	5	Interactive
Apathetic	1	2	3	4	5	Responsive

3.4.1.2. Implementation

After the discussions, participants were asked to complete the Godspeed Questionnaire via Google Forms, to measure their perception of voice assistants. After the experiment, users were then given the same questionnaire to complete to see if

their perception had changed. Each time they were asked to read information about the questionnaire and provide their consent (as seen in Appendix J).

3.4.2. The Checklist for Trust between People and Automation

The Checklist for Trust between People and Automation (Jian et al,. 2000) is a tool made up of twelve statements with Likert scales ranging from 1 to 7. Existing tools at the time tried to examine trust, but did not differentiate between human-human trust, human-machine trust as well as general trust, thus the questionnaire was created to accurately measure the human-machine trust that exists between people and automated systems. To ensure an overall score can be calculated, the first five statements are reversed for a higher level of trust to be represented by a higher score. This means an average score closer to 7 will represent a higher level of trust.

3.4.2.1. Sample

A sample is shown below, the full questionnaire is available in Appendix B.

The system is deceptive							
Not at all	1	2	3	4	5	6	7	Extremely
I am suspicious of the system's intent, action, or outputs							
Not at all	1	2	3	4	5	6	7	Extremely
The system is reliable							
Not at all	1	2	3	4	5	6	7	Extremely

3.4.2.2. Implementation

Users were asked to complete The Checklist for Trust between People and Automation via Google Forms, as well as to read information about the questionnaire and sign their consent.

Frequent users were asked to continue using the device like normal and infrequent users were asked to integrate it into their lives and use it too. Both groups were not given a specific set of tasks, but were given suggestions concerned with achieving a goal (such as setting alarms, finding out a fact, asking about a celebrity) and concerned with fun (such as quizzes, singing, jokes). The reason for this is that the

participants would interact with the voice assistant in their own natural manner, rather than having their actions controlled.

Additionally, both sets of participants were asked to carry out these activities by changing the humanlike-nature of the system through personalisation of the voice. The reason for this is that only one participant had changed the voice in the past (and could not remember anything about it), everyone else had never considered it. They were asked to select a voice that they felt comfortable interacting with.

Both groups were asked to do this for a week. At the end of this, the groups completed the same questionnaire and the results were then compared with the results from their first instance of completing it. This would help understand if trust could be affected by modifying the voice assistant's voice.

3.4.3. Subjective Assessment of Speech System Interfaces

The SASSI (Hone & Graham, 2000) is a questionnaire made of thirty-four statements with Likert scales ranging from Strongly Disagree to Strongly Agree and is used to measure speech-system usability. These statements are organised into six sections: System Response Accuracy, Likeability, Cognitive Demand, Annoyance, Habitability and Speed.

- System Response Accuracy refers to the system accurately recognising the user's command and as well as the system's resulting response
- Likeability refers to the how much the user likes the enjoys using the system
- Cognitive Demand refers to the effort required to use the system as well as the feelings that appear from it
- Annoyance refers to how annoyed users feel when using the system
- Habitability refers to how well the user knows how to use the system
- Speed refer to how quickly the system acts (Hone & Graham, 2000)

It is a recommended tool (Lewis, 2016) in the evaluation of speech based system with its high usage (Lewis & Sauro, 2020) and its large coverage of UX dimensions (Kocabalil et al., 2018). Just like with the previous questionnaire, a similar change was made: all negatively formed statements have been inverted to keep scores consistent in each section. All scores from "annoyance" were reversed to make the larger number the more positive score to align with the other sections.

3.4.3.1. Sample

A sample is shown below, the full questionnaire is available in Appendix C.

The system is accurate								
Strongly Disagree	1	2	3	4	5	6	7	Strongly Agree
I felt confident using the system								
Strongly Disagree	1	2	3	4	5	6	7	Strongly Agree

I always knew what to say to the system

Strongly Disagree	1	2	3	4	5	6	7	Strongly Agree

3.4.3.2. Implementation

To help explore whether or not there is a relationship between trust and speech-system usability, users were asked to complete the SASSI via Google Forms. They were asked to read information about the questionnaire and sign their consent. These results were then compared with the trust scores to understand if trust correlates with usability.

3.4.4. Qualitative Research

3.4.4.1. Focus Group

A focus group is a technique that involves a moderator and a group of participants to explore thoughts and feelings around a certain topic. They encourage discussion amongst the party as one member may bring up an idea that the others might not have considered and this could possibly lead to the rest of the group relaying their thoughts in relation to that initial topic. The researcher may choose to use a list of set questions; however, a semi-structured approach can be taken to further explore participants' statements. This approach can also make the group feel more comfortable by creating a more casual setting.

3.4.4.2. Interviews

An interview is a discussion in which the researcher will ask a user questions on a specific topic. Where a focus group has multiple members, an interview will often consist of just one individual. As mentioned with focus groups, the interview can be semi-structured to create a more informal setting and help to relax the participant.

3.4.4.3. Implementation

To explore perception of the users, focus groups and interviews were carried out depending on the participants' availability. If several individuals were available at the same time, they were organised into a focus group, and if a participant's schedule did not work out, an interview was organised. These took place online via Skype. For the frequent users group, one focus group of four people and one focus group of two people were carried out. Both methods were semi-structured to create a more open discussion and allow for further exploration based on the respondents' answers. For the infrequent users group, two focus groups, each with two people, were carried out as well two interviews each with one individual. Again, both methods were semi-structured for a more casual conversation. Both groups of users were asked questions on their experience with voice assistants in the areas of "general usage", "trust" and "social presence".

Prior to taking part in the focus group/interview, participants received an information sheet (as seen in Appendix F) as well as a consent form (as seen in Appendix G) via

email. The forms were signed and returned, and discussions could begin. Questions were generated from literature review research and are as follows:

Frequent users were asked the following questions on their general usage of voice assistants:

- “What is your general opinion of voice assistants?”
- “How long would you say you have been using voice assistants for?”
- “What voice assistants have you used?”
- “Is there a reason why you would choose one over the others?”
- “What made you start?”
- “What typical things do you do with them?”
- “Is there anything you would refuse to use them for?”

Infrequent users were asked the following questions on their general usage of voice assistants:

- “What is your general opinion of voice assistants?”
- “What voice assistants have you used?”
- “What turned you away from using them/it?”
- “What typical things have you tried to do with them?”
- “Is there anything you would refuse to use them for?”

Further questions were then asked based off the answers given and participants discussed amongst each other at certain points.

Both sets of users were asked the following questions on trust in voice assistants:

- “Is trust something you have ever considered when using your voice assistant?”
- “Is privacy a concern?”
- “Are you concerned at all with the organisation respective to your voice assistant?”
- “Do you trust the system to carry out its own capabilities”

Both sets of users were asked the following questions on trust in voice assistants:

- “Have you ever considered the humanlike-nature of your voice assistant?”

- “Have you ever considered the voice assistant has a personality?”
- “Have you ever tried to personalise the voice/accent?”
- “What is your opinion on voice assistants being adapted into more humanoid shapes, such as robots?”

3.4.5. Data Analysis

3.4.5.1. Statistical Analysis

To understand the difference between two sets of data, statistical analysis can be carried out to assist the researcher in comprehending the data better. One example that will be used in this research is a t-Test, which can tell the researcher if the difference between groups of data is significant. One set of data may score higher than the one it is being compared to, however the difference between the two may not be large enough to decide that there that is a certain disparity.

3.4.5.2. Thematic Analysis

Thematic analysis is an analysis of qualitative data such as focus group transcripts. The text is examined to determine the topics and patterns that repeatedly appear. There are two main approaches that can be taken in thematic analysis: inductive and deductive. Inductive refers to using the collected data to generate the themes whereas deductive refers to approaching the data having already considering themes beforehand. A common process of this analysis is: Familiarise, coding, generate themes, review themes, define and name themes, writing up (Caulfield, 2019).

- Familiarise: read through and fully understand the data that has been collected
- Coding: read through the data and organise sections of it into groups
- Generate themes: create themes that can combine multiple codes
- Review themes: look at the data again to make sure the themes are appropriate, if not then make new ones or split them up further and combine again until the themes are completely relevant
- Define and name themes: label each theme by creating an understandable name for each one, and give each theme a definition
- Writing up: create the data analysis (Caulfield, 2019).

3.5. Summary

This chapter has explained the methodology and approach taken to assist in exploring the main research question, as well as other approaches that could have been used. It described the methods and tools used to achieve the aims of the research and explained how each method and tool was used in the research.

4. Analysis of Gathered Data

This chapter will provide the data gathered from the study as well as appropriate analysis for each set of results.

4.1. Initial Focus Groups/Interviews

Both sets of users were organised into interviews and focus groups to provide descriptions and insight into what each group had to say on voice assistants.

Table 2

Focus Groups/Interviews

Discussion	**Participants Involved**	**Length**
Focus Group 1	P2, P3, P4, P5	35 minutes
Focus Group 2	P1, P6	25 minutes
Focus Group 3	P7, P10	25 minutes
Focus Group 4	P11, P12	20 minutes
Interview 1	P8	20 minutes
Interview 2	P9	20 minutes

Note: P[1-12] = Participant[1-12]

Thematic analysis was used to identify common themes from the discussions. The discussions were revolved around general usage, trust and social presence, so themes naturally formed around those and the questions asked. Having completed a literature review, a deductive approach was taken to this analysis as topics that could come up in the discussions were considered beforehand. The theme names were considered previously, however the results of the discussions helped shape these.

As the discussions were separated into frequent and infrequent users, different themes emerged from both.

The first step was to understand the data that had been collected. The participant recordings were then read through, and data was organised into codes. These codes were then used to create themes that could combine multiple pieces of information. These themes were reviewed until the most appropriate themes were created. An example is shown below:

Quotes

"P3: I would never use a voice assistant to look up programming information, as it doesn't provide the correct answers, whereas I would always set timers with my voice assistant"

"P4: One thing I use mine for is to change the lights in my house and something I would not use it for is to make calls, as I would want to make sure I got it right first time"

Code

Easy tasks

Difficult tasks

Theme

Simple Tasks Only

4.1.1. Frequent Users Themes

4.1.1.1. Simple Tasks Only

Frequent users spoke about how their main reason for using a voice assistant was due to completing tasks and achieving goals. Their main tasks included setting timers, asking for definitions, asking how to spell words and using it when in a hands-free situation (in the kitchen or in the car). Two users mentioned having it connected to their home's lights and would use the device to control these. Users generally thought these were simple and easy to complete.

P1: "One of the main things I use Alexa for is to play music in the car, and it does this well most of the time"

However, there were some goal-orientated tasks that frequent users would refuse to use a voice assistant for, and these were instances where the individual thought could be difficult for the voice assistant to carry out, and they would want to get it right the first time. Frequent users believed that doing an important task poorly could be problematic.

P2: "If I was to send a work-related email, or make a phone call to a colleague via my voice assistant and it did not work the first time, it might appear as unprofessional. That is the kind of task that I want to get correct straight away, and not have to attempt twice"

4.1.1.2. Voice Assistants are Appealing

The group was asked questions on their opinions surrounding social presence. While frequent users did consider the downside of social presence (such as creepiness), it was generally agreed that the prospect of the voice assistant becoming more humanlike was an exciting one. Referring to science fiction and superhero movies, frequent users showed excitement.

P6: "It's like having my own version of C3-P0 or J.A.R.V.I.S. from the Avengers films."

There was some consideration to the negativity that could appear from science fiction (such as voice based technology taking over the home) becoming true, however the users showed little concern for this realistically ever happening.

P2: "Of course, there is the fear that if these go too far then it could be dangerous. If every appliance in my home is connected to the voice assistant, and if someone with malicious intent is planning to control me, could they hack into my voice assistant and control my devices? I know that's highly unlikely, but some people could be worried that what they see in movies could become real"

4.1.1.3. Individual Knowledge

Frequent users showed a personal understanding into how voice assistants work, and how they should be interacted with. There was little frustration as they knew that some tasks could be carried out and that some could not.

P3: "From experience of using the voice assistant, I know there are some things it cannot do, and that does not really annoy me. I know the functionality does not exist yet so I do not get frustrated"

They also stated that they will speak clearer when interacting with the device, as they understand that the assistant will have trouble picking up on certain aspects of speech such as accents or specific words.

P4: "My accent is a bit thicker, so the device has trouble picking up on certain things that I say, but I know now that I just have to adjust how I say a few words which is easy enough"

4.1.1.4. Trust: Accepting

When asked about trust, frequent users mentioned slight concern into the device always listening, however they generally accepted that the companies who manufacture the voice assistants already have their data due to years of using other services, so while they acknowledge the issue of security, they understand that it is something they must recognise.

P6: "I've been using Google's services for years, so I know that they already contain a lot of my data anyway. I have Google devices in most rooms of my flat, so I've kind of agreed to let them listen to me anyway. Even if it's still wrong I come to accept that it's the trade-off for using many of their free services"

4.1.2. Infrequent Users

4.1.2.1. Simple Tasks Only

Infrequent users also showed a preference for completing simple tasks. They spoke about how in their little experience of using voice assistants they had only ever used it to carry out basic tasks like asking a question or playing a song. Infrequent users also stated that they would refuse to use the voice assistant for anything that required personal information such as medical related tasks or shopping.

P9: "I only ever use it for the radio, I would never use it for something personal like health"

P10: "I would not purchase something with the voice assistant. I'd like to see that the purchase has gone through and not being able to visually know that the payment has been successful is worrying to me"

4.1.2.2. Voice Assistants Are Intrusive

Infrequent users generally showed concern towards the untrustworthiness of the voice assistants in terms of both functionality and privacy. As stated before, only basic tasks were carried out, however, frustration towards these tasks being answered incorrectly caused most of the infrequent users to stop using the device all together. Some of the group also worried that the device is always listening even when it does not appear to be.

P7: "I turn it off at the wall when I am not using it just in case it listens to me"

Many brought up the fact that if it is listening for a keyword, then it must always be listening for the keyword to be heard. The term "uncanny valley" was brought up several times. When presented with the idea of the voice assistant appearing as more of a humanoid shape, infrequent users felt that this would be too intrusive and add to how creepy it already feels by listening.

P9: “I am already worried that the device is listening to me, if it was humanoid then I think I would be worried that it was watching me too. I would probably be happy with just the voice changing, but not it’s appearance”

4.1.2.3. Voice Assistants Are Unnecessary

Some infrequent users spoke about how they believed the voice assistant did not offer any value to them, that it would be much easier to just manually use their phone to retrieve the information. They had used one in the past, however integrating it into their routine was not worth the effort.

P11: “If I am completing a task, I think I can probably do it quicker and more successfully than if I was to use a voice assistant for it.”

P8: “I guess I have not be able to find a place for it in my daily routines, so I just can’t see how it would be of use to me”

4.2. Godspeed Questionnaire

Data from the Godspeed Questionnaire can be seen in tables 3, 4, 5 and 6, and figures 1, 2, 3 and 4.

Table 3

Godspeed Questionnaire (Frequent Users compared with Infrequent Users)

t-Test (Two-Sample Assuming Equal Variances) (Significance Level = 0.05)

	Frequent User Mean	Infrequent User Mean	P-Value
Anthropomorphism	2.70	2.63	= 0.71
Animacy	3.61	2.47	< 0.01
Likeability	4.23	3.07	< 0.0001
Perceived Intelligence	3.73	3.27	= 0.20
Perceived Safety	3.61	3.45	= 0.79
Overall (Before)	3.58	2.98	= 0.09
Overall (After)	3.66	2.95	= 0.03

Note: This table shows information about the significance in differences between the groups in individual perception categories

Table 4

Godspeed Questionnaire (Frequent users (before) compared with frequent users (after) | Infrequent users (before) compared with infrequent users (after))

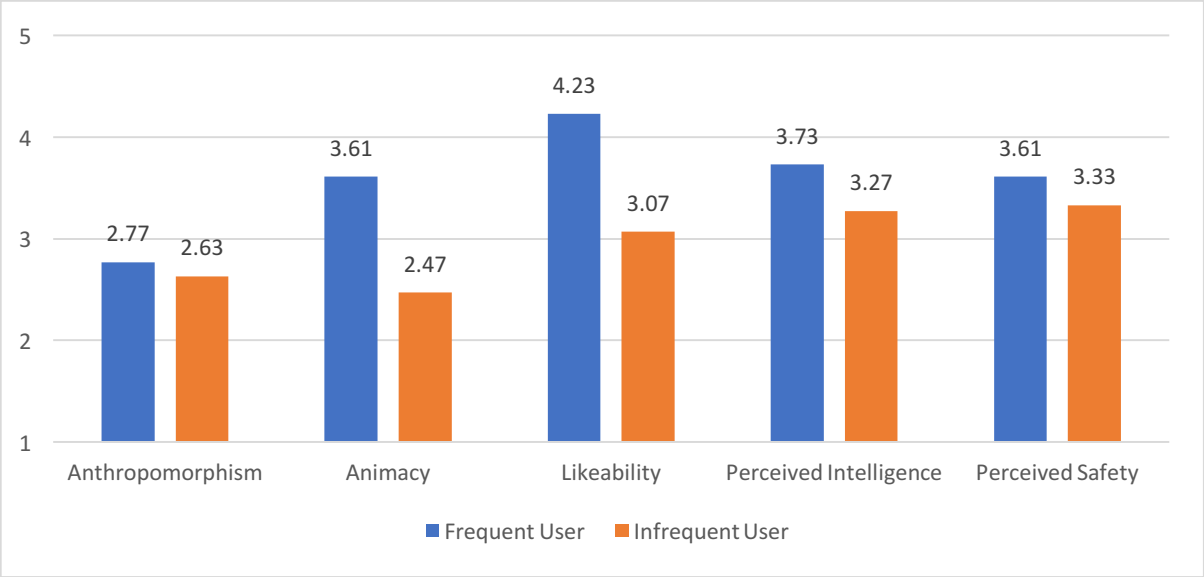
t-Test (Paired Two Sample for Means) (Significance Level = 0.05)

	Mean (Before)	**Mean (After)**	**P-Value**
Frequent User	3.58	3.66	= 0.36
Infrequent User	2.98	2.95	= 0.87

Note: This table shows information about the significance in differences of each group

Figure 1

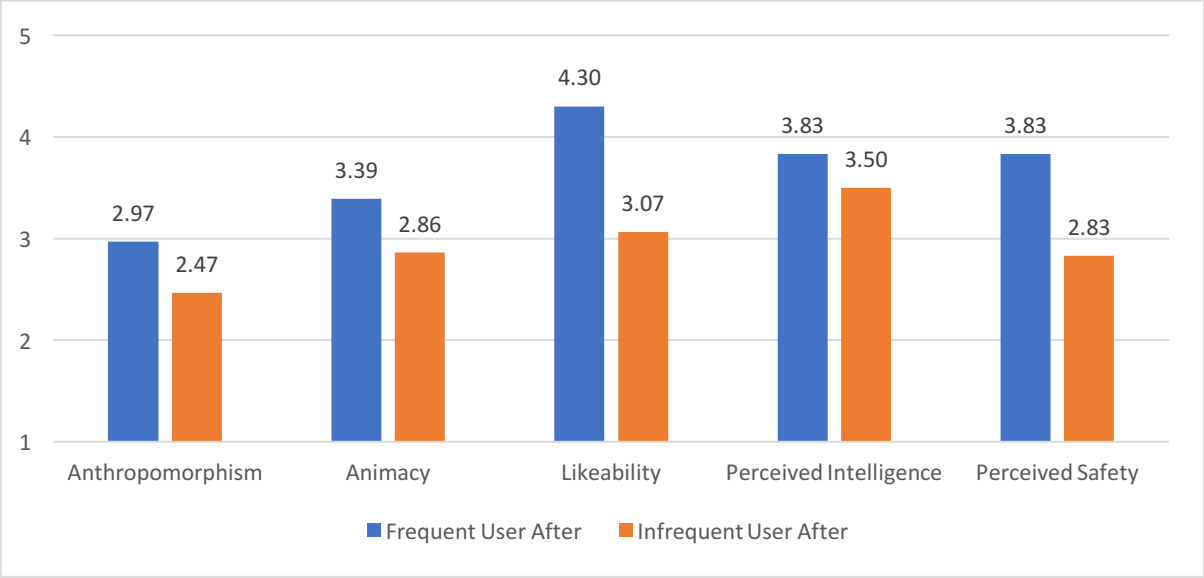
Godspeed Questionnaire - Frequent Users/Infrequent Users (Before)



Note: This graph compares the "before" perception scores of frequent users and infrequent users

Figure 2

Godspeed Questionnaire - Frequent Users/Infrequent Users (After)



Note: This graph compares the "after" perception scores of frequent users and infrequent users

Table 5

Godspeed Questionnaire - Frequent User Results (Before/After)

	Frequent User Before	Frequent User After	Difference
Anthropomorphism	2.70	2.97	+0.27
Animacy	3.61	3.39	-0.22
Likeability	4.23	4.30	+0.07
Perceived Intelligence	3.73	3.83	+0.10
Perceived Safety	3.61	3.83	+0.22
Overall mean	3.58	3.66	+0.08

Note: This table shows information on how perception has changed in frequent users

Table 6

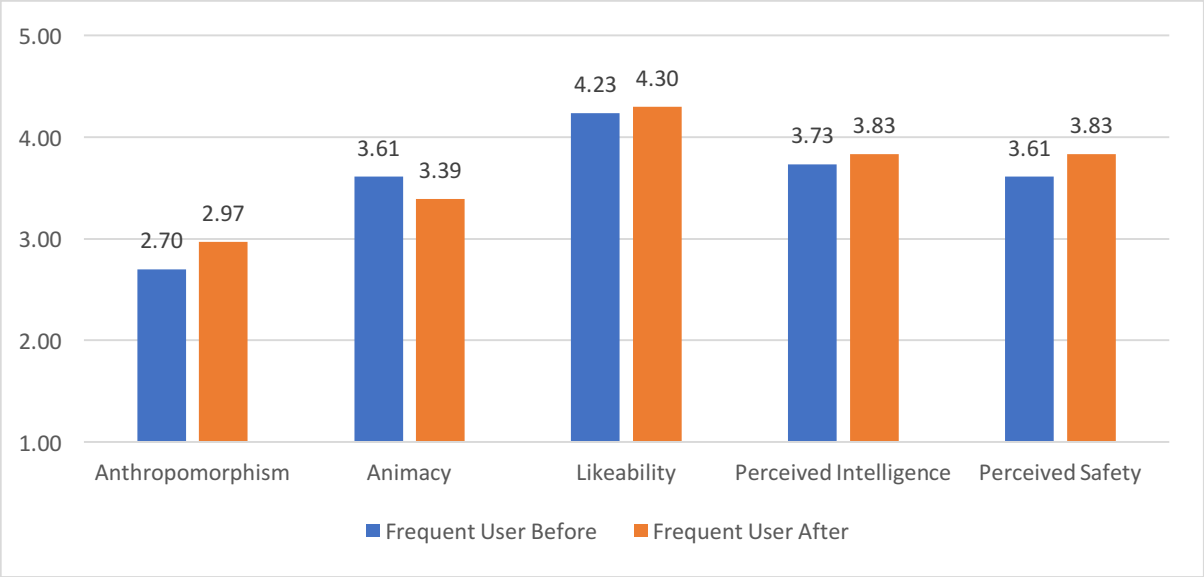
Godspeed Questionnaire - Infrequent User Results (Before/After)

	Infrequent User Before	Infrequent User After	Difference
Anthropomorphism	2.63	2.47	-0.16
Animacy	2.47	2.86	+0.39
Likeability	3.07	3.07	+0.00
Perceived Intelligence	3.27	3.50	+0.23
Perceived Safety	3.45	2.83	-0.62
Overall mean	2.98	2.95	-0.03

Note: This table shows information on how perception has changed in infrequent users

Figure 3

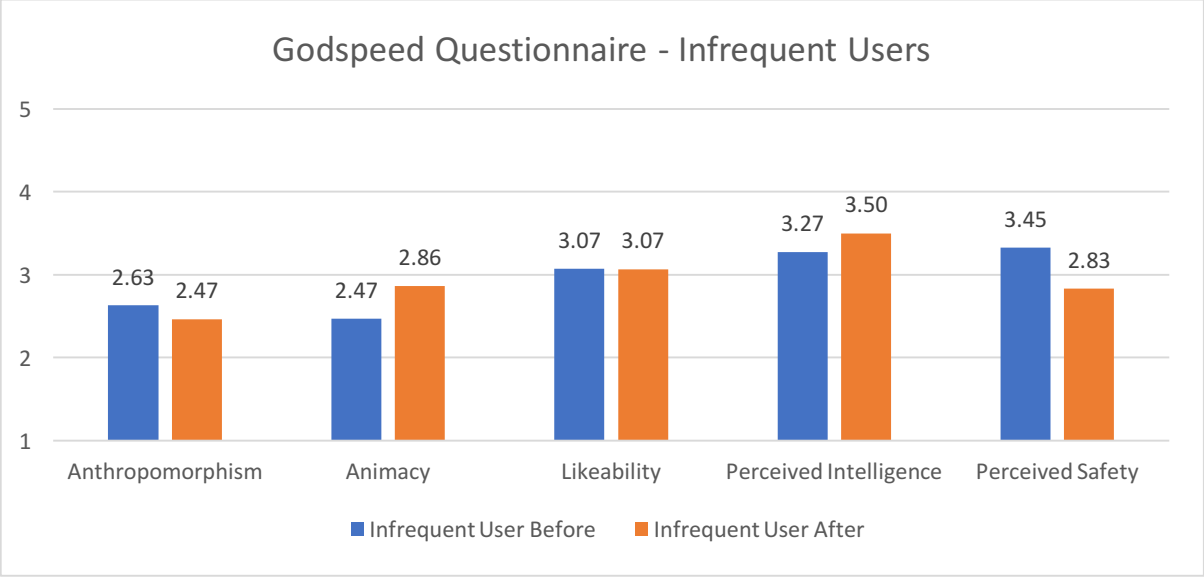
Godspeed Questionnaire - Frequent Users (Before/After)



Note: This graph compares the "before" and "after" perception scores of frequent users

Figure 4

Godspeed Questionnaire - Infrequent Users (Before/After)



Note: This graph compares the "before" and "after" perception scores of infrequent users

4.3. The Checklist for Trust between People and Automation

Data for The Checklist for Trust between People and Automation can be seen in tables 7, 8, 9 and 10, and figures 5, 6, 7, 8, 9, 10, 11 and 12.

Table 7

Checklist for Trust between People and Automation (Frequent Users compared with Infrequent Users)

t-Test (Two-Sample Assuming Equal Variances) (Significance Level = 0.05)

	Frequent User Mean	Infrequent User Mean	P-Value
Before	4.95	3.89	= 0.02
After	5.20	3.81	= 0.02

Note: This table shows information about the significance in difference of users' trust before and after

Table 8

Checklist for Trust between People and Automation (Frequent users (before) compared with frequent users (after) | Infrequent users (before) compared with infrequent users (after))

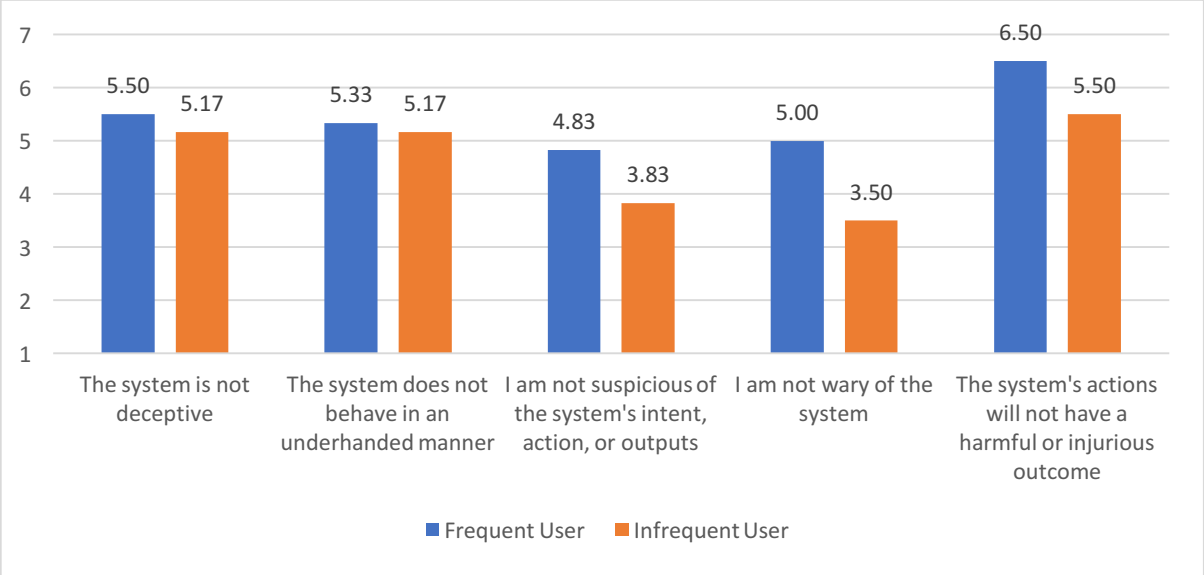
t-Test (Paired Two Sample for Means) (Significance Level = 0.05)

	Mean (Before)	Mean (After)	P-Value
Frequent User	4.95	5.20	= 0.32
Infrequent User	3.89	3.81	= 0.86

Note: This table shows information about the significance in differences of trust in each group

Figure 5

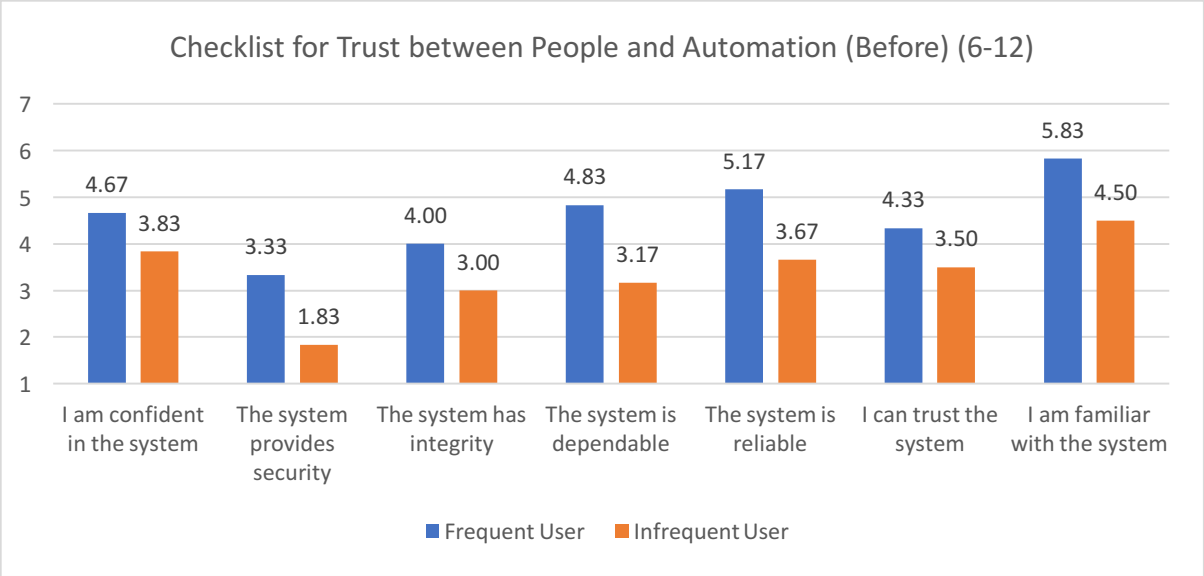
Checklist for Trust between People and Automation - Frequent Users/Infrequent Users (Before) (1-5)



Note: This graph compares the "before" trust scores for frequent users and infrequent users for statements 1 to 5

Figure 6

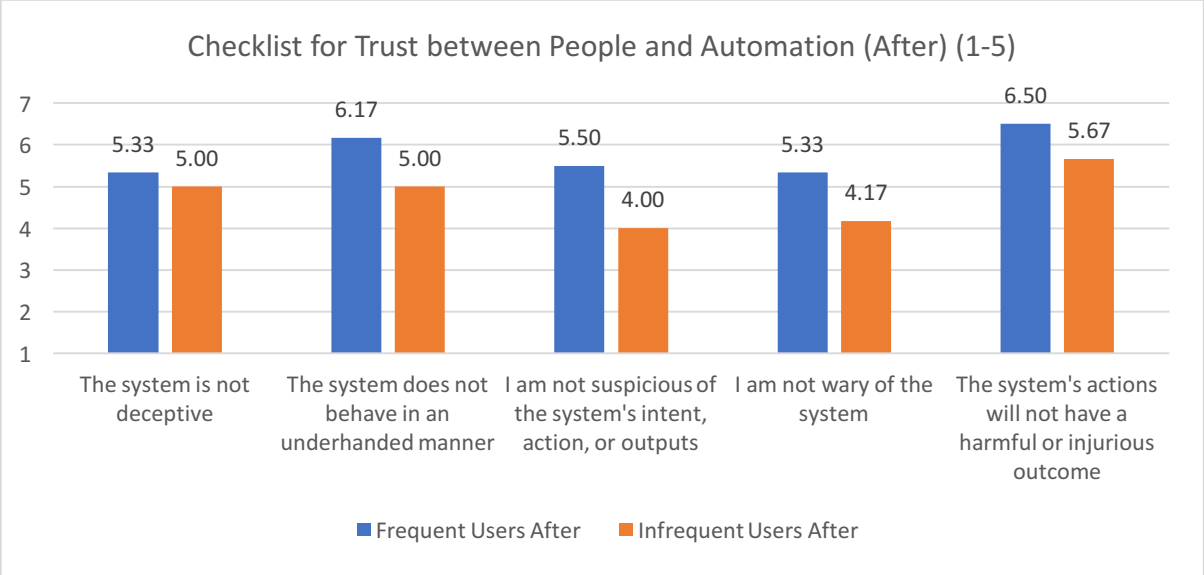
Checklist for Trust between People and Automation - Frequent Users/Infrequent Users (Before) (6-12)



Note: This graph compares the "before" trust scores for frequent users and infrequent users for statements 6 to 12

Figure 7

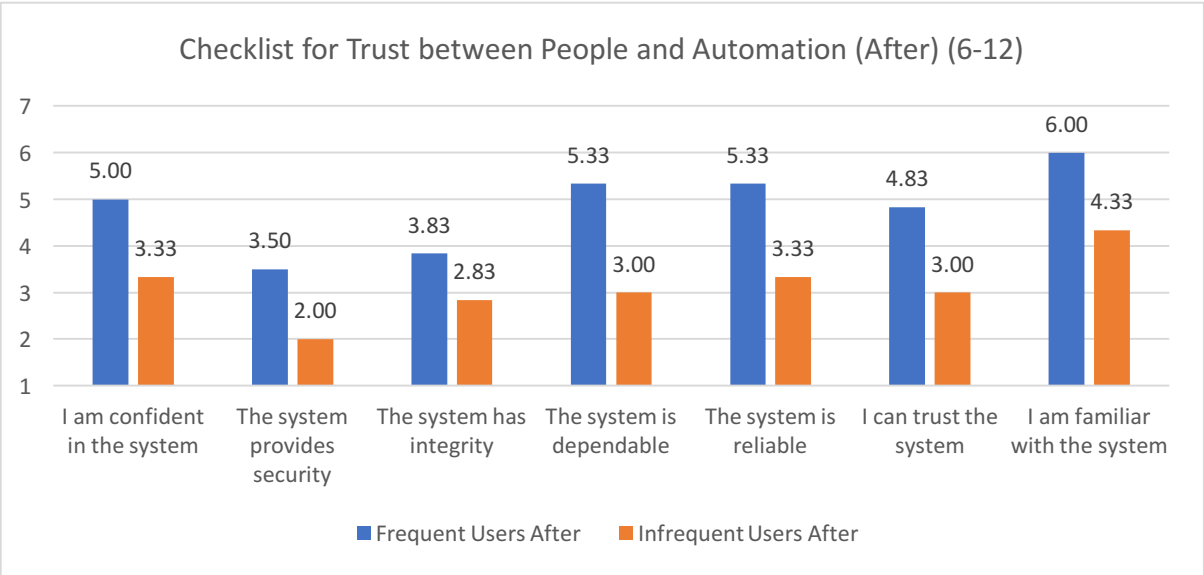
Checklist for Trust between People and Automation - Frequent Users/Infrequent Users (After) (1-5)



Note: This graph compares the "after" trust scores for frequent users and infrequent users for statements 1 to 5

Figure 8

Checklist for Trust between People and Automation - Frequent Users/Infrequent Users (After) (6-12)



Note: This graph compares the "after" trust scores for frequent users and infrequent users for statements 6 to 12

Table 9

Checklist for Trust between People and Automation - Frequent Users Results (Before/After)

Participant	Before	After	Difference
p1	5.33	6.42	**+1.08**
p2	4.08	3.67	**-0.42**
p3	4.92	4.75	**-0.17**
p4	4.92	4.92	**0.00**
p5	5.17	5.50	**+0.33**
p6	5.25	5.92	**+0.67**

Note: This table shows information on how trust has changed in frequent users

Table 10

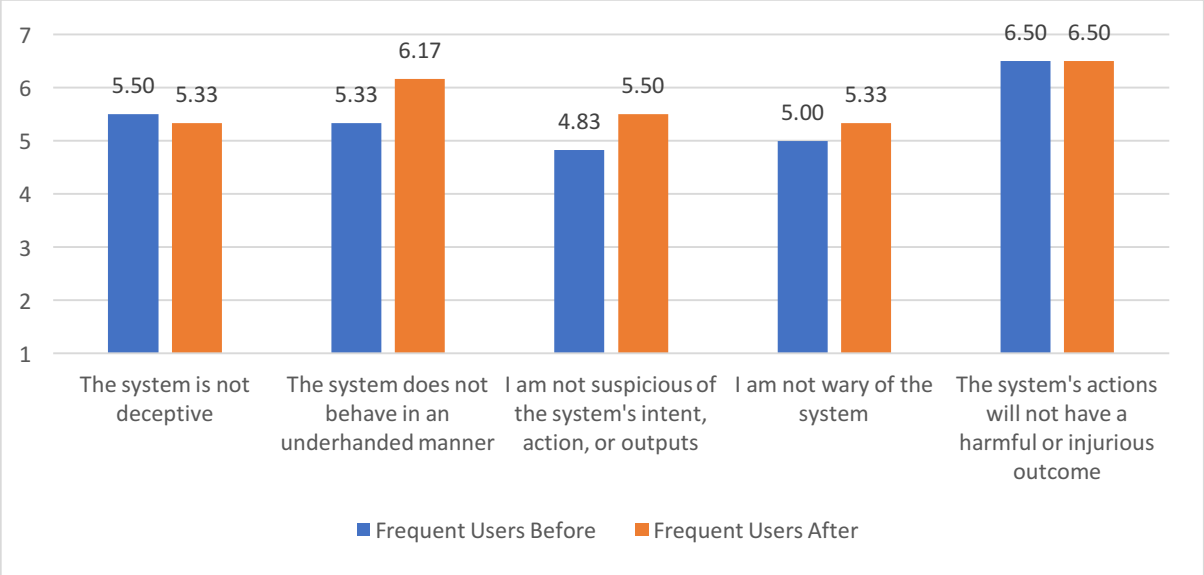
Checklist for Trust between People and Automation - Infrequent Users Results (Before/After)

Participant	Before	After	Difference
p7	3.92	4.00	**+0.08**
p8	4.17	4.58	**+0.41**
p9	3.33	4.42	**+1.09**
p10	3.83	2.67	**-1.16**
p11	2.75	3.42	**+0.67**
p12	5.33	3.75	**-1.58**

Note: This table shows information on how trust has changed in infrequent users

Figure 9

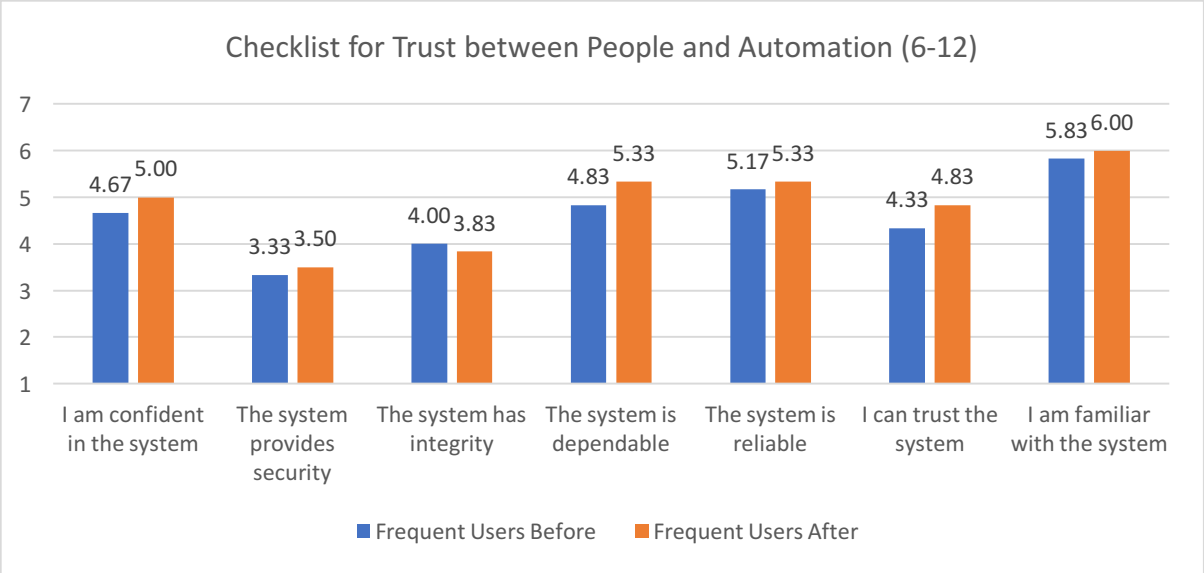
Checklist for Trust between People and Automation - Frequent Users (Before/After) (1-5)



Note: This graph compares the "before" and "after" trust scores for frequent users for statements 1 to 5

Figure 10

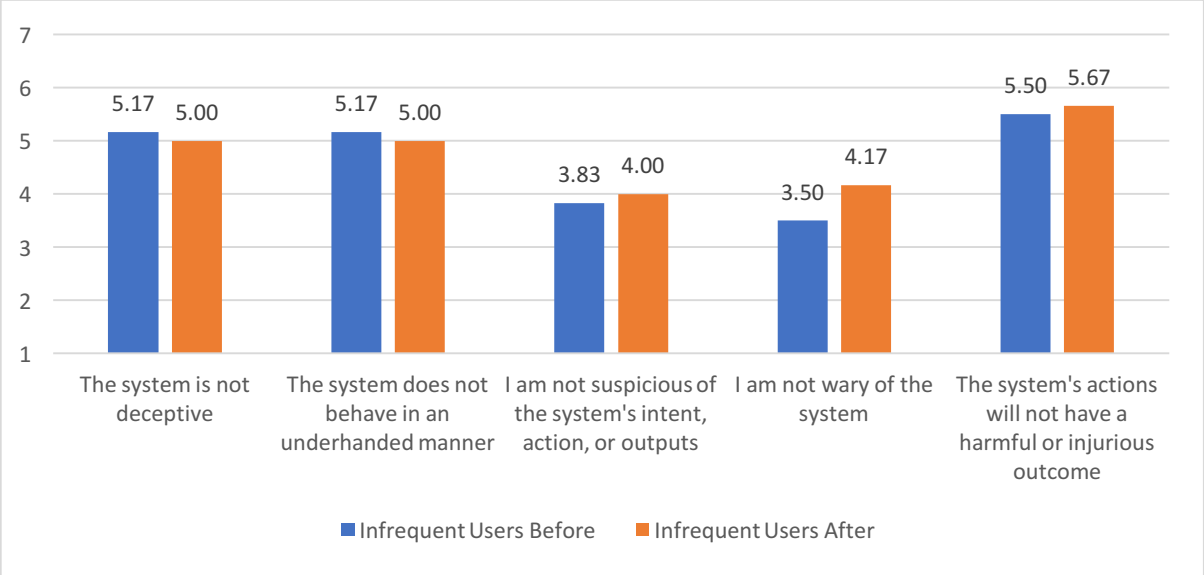
Checklist for Trust between People and Automation - Frequent Users (Before/After) (6-12)



Note: This graph compares the "before" and "after" trust scores for frequent users for statements 6 to 12

Figure 11

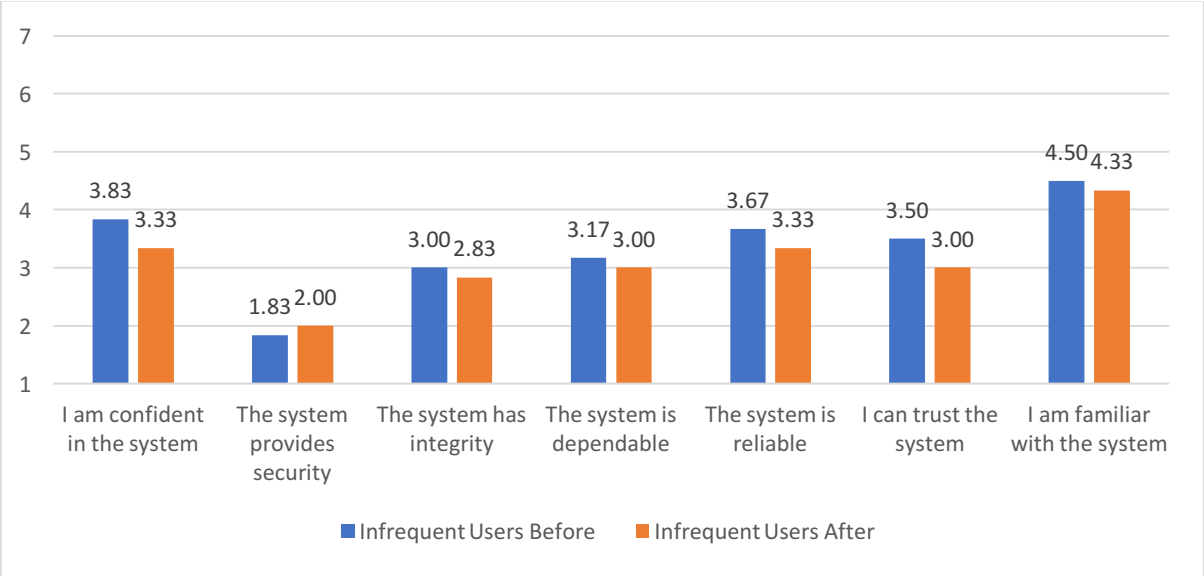
Checklist for Trust between People and Automation - Infrequent Users (Before/After) (1-5)



Note: This graph compares the "before" and "after" trust scores for infrequent users for statements 1 to 5

Figure 12

Checklist for Trust between People and Automation - Infrequent Users (Before/After) (6-12)



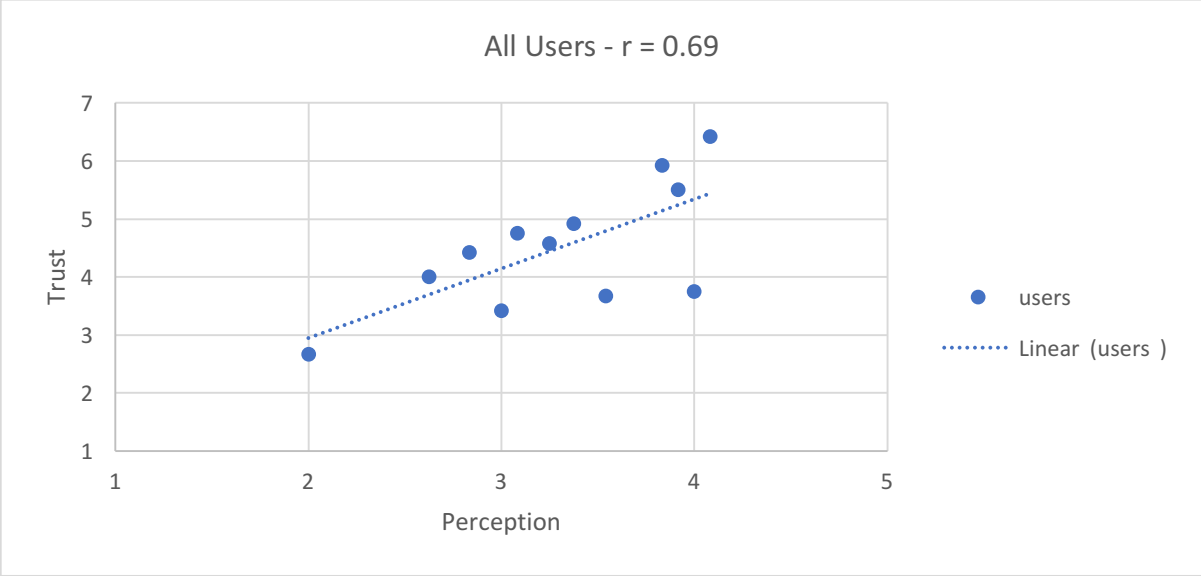
Note: This graph compares the "before" and "after" trust scores for infrequent users for statements 6 to 12

4.4. Godspeed Questionnaire/The Checklist for Trust between People and Automation

Correlations between trust and perception can be seen in figures 13, 14 and 15.

Figure 13

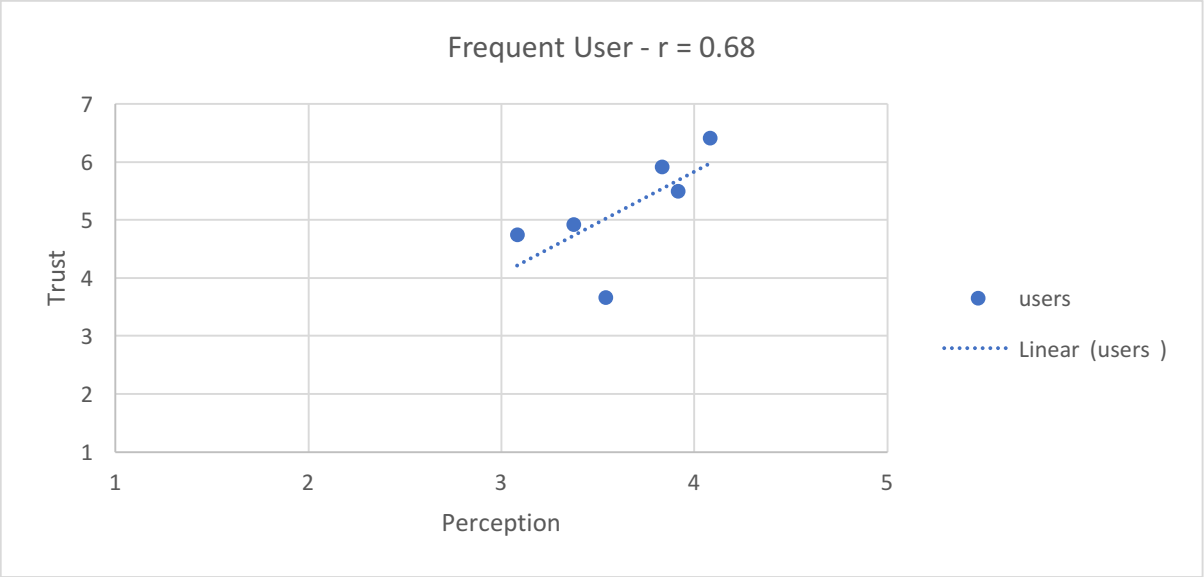
Trust and Perception



Note: This graph shows the correlation of trust and perception in frequent users and infrequent users

Figure 14

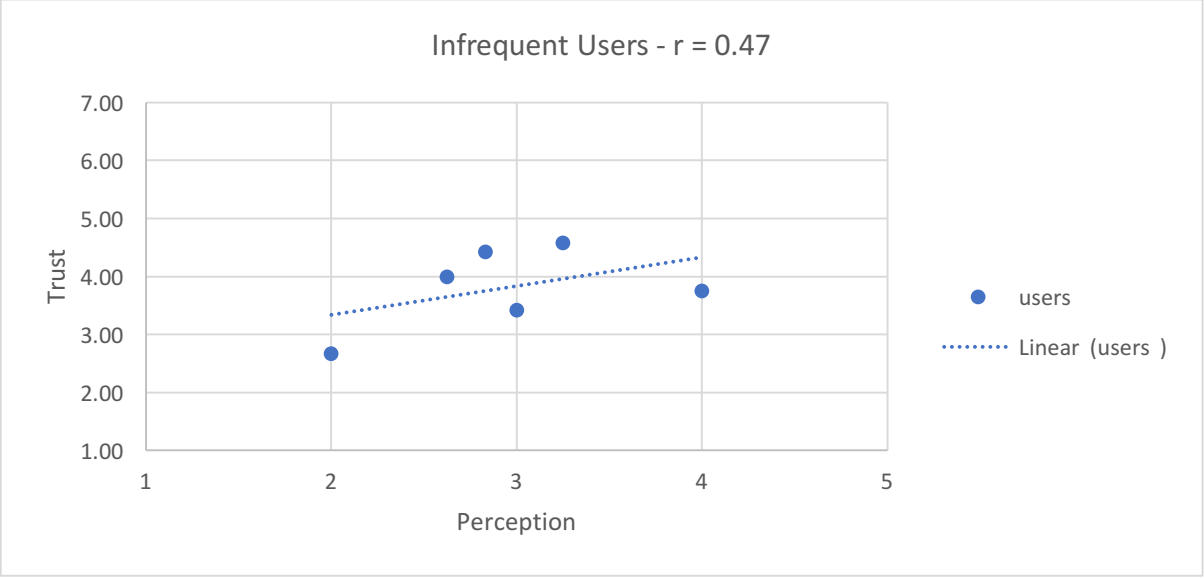
Trust and Perception - Frequent Users



Note: This graph shows the correlation of trust and perception in frequent users

Figure 15

Trust and Perception - Infrequent Users



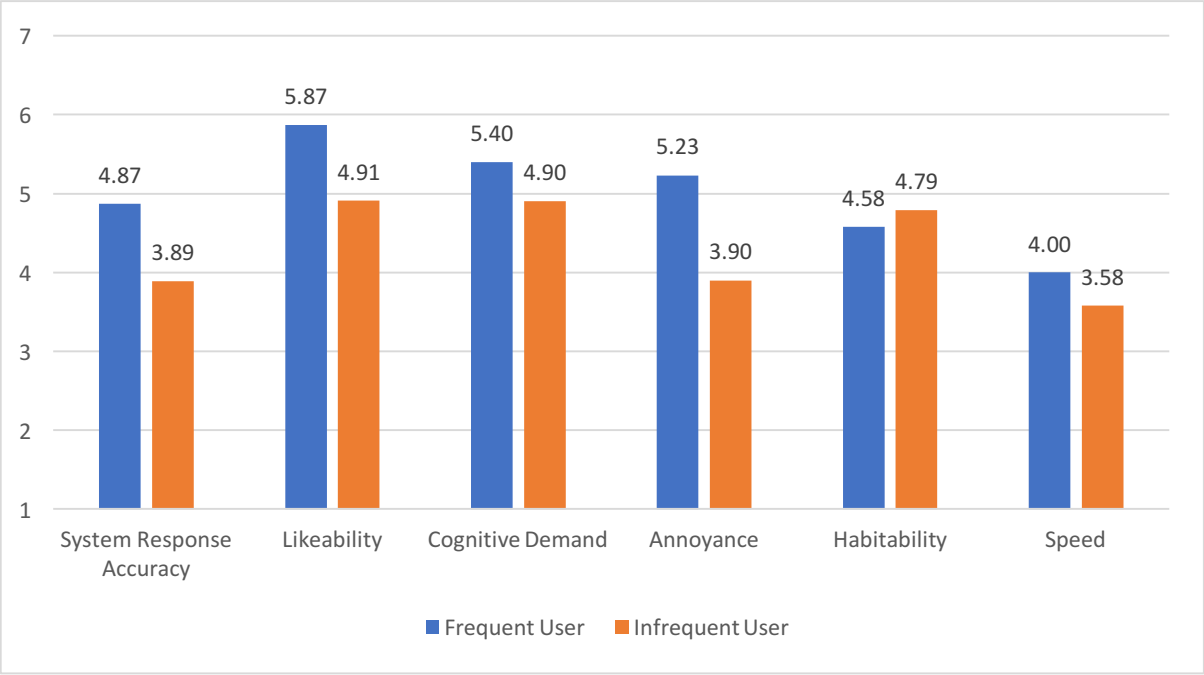
Note: This graph shows the correlation of trust and perception in infrequent users

4.5. Subjective Assessment for Speech System Interfaces (SASSI)

Data for the SASSI can be seen in figure 16.

Figure 16

SASSI - Frequent Users/Infrequent Users



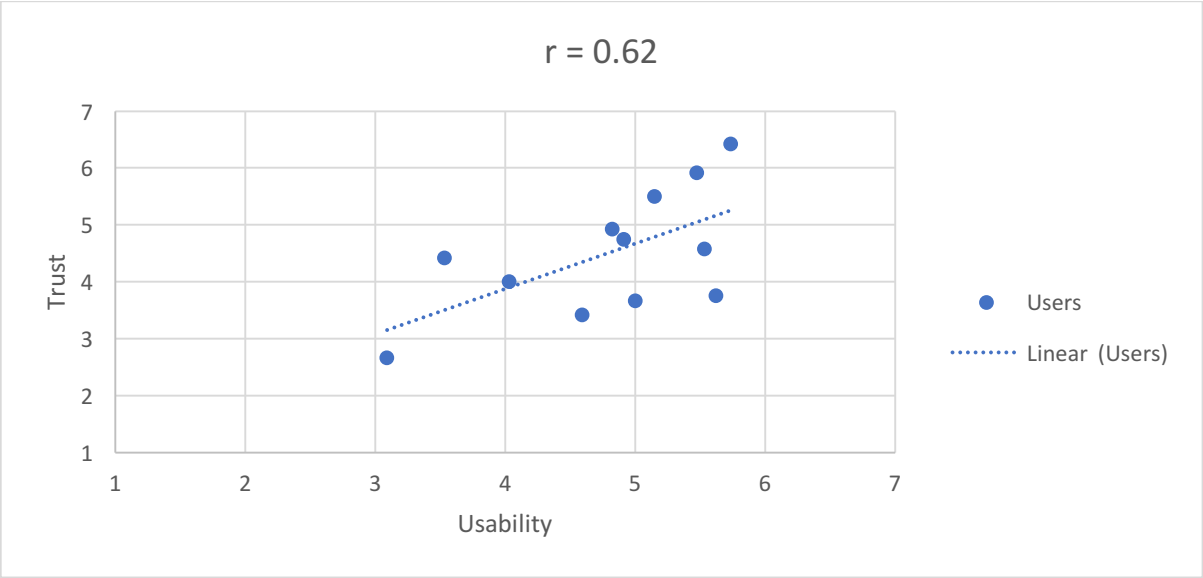
Note: This graph compares the speech-system usability scores for frequent users and infrequent users

4.6. SASSI/The Checklist for Trust between People and Automation

The correlation between overall speech-system usability and trust can be seen in figure 17, and the correlation between the individual factors of the SASSI can be seen in figures 18, 19, 20, 21, 22 and 23.

Figure 17

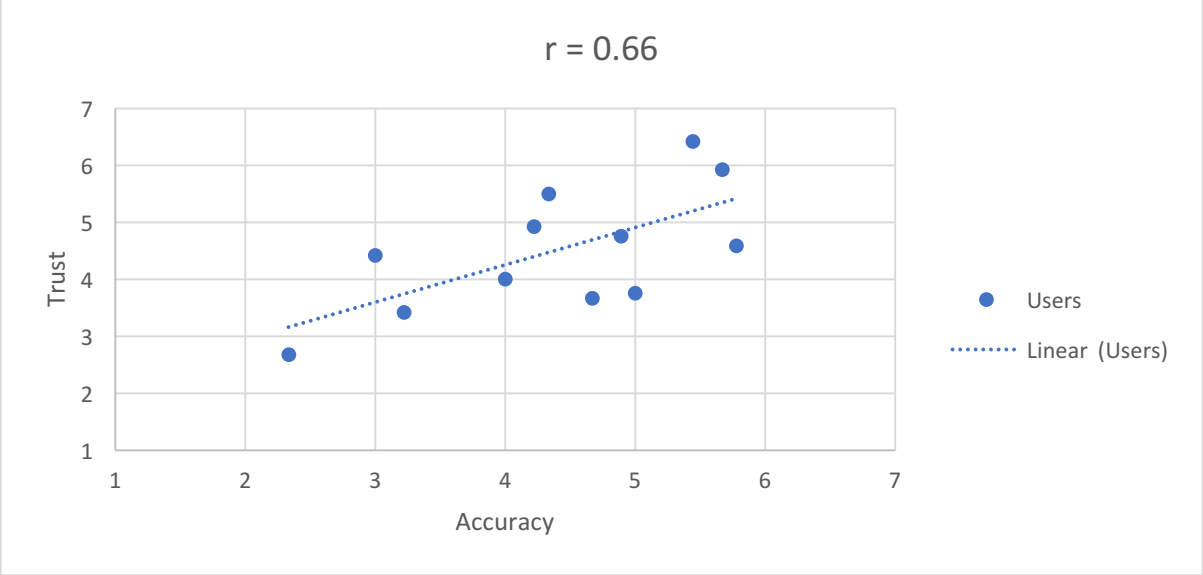
Speech-System Usability and Trust



Note: This graph shows the correlation of trust and speech-system usability

Figure 18

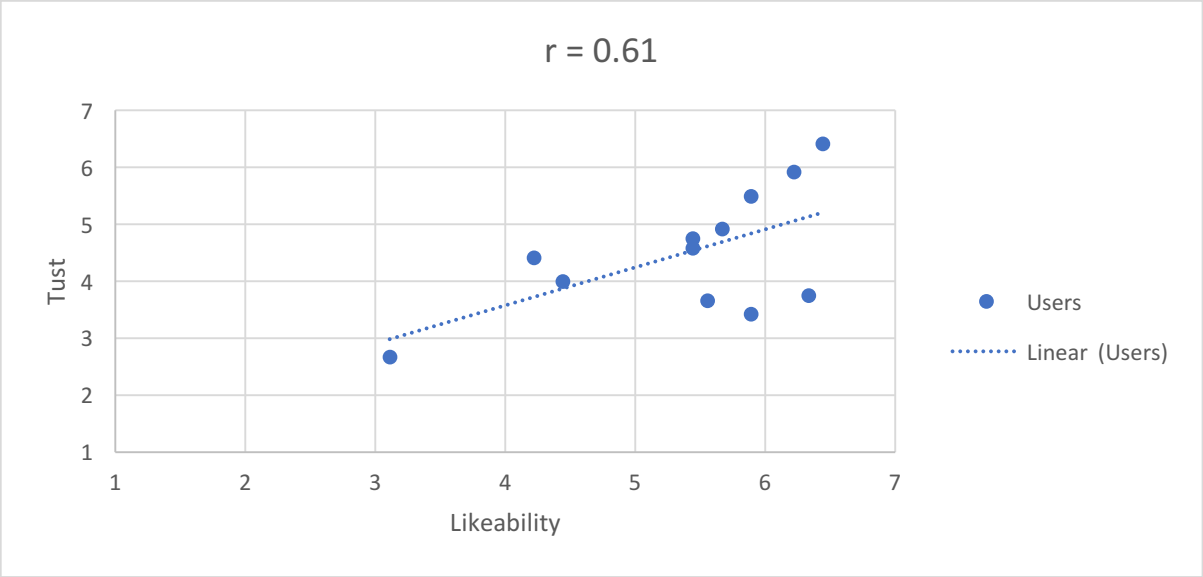
Accuracy and Trust



Note: This graph shows the correlation of trust and system response accuracy

Figure 19

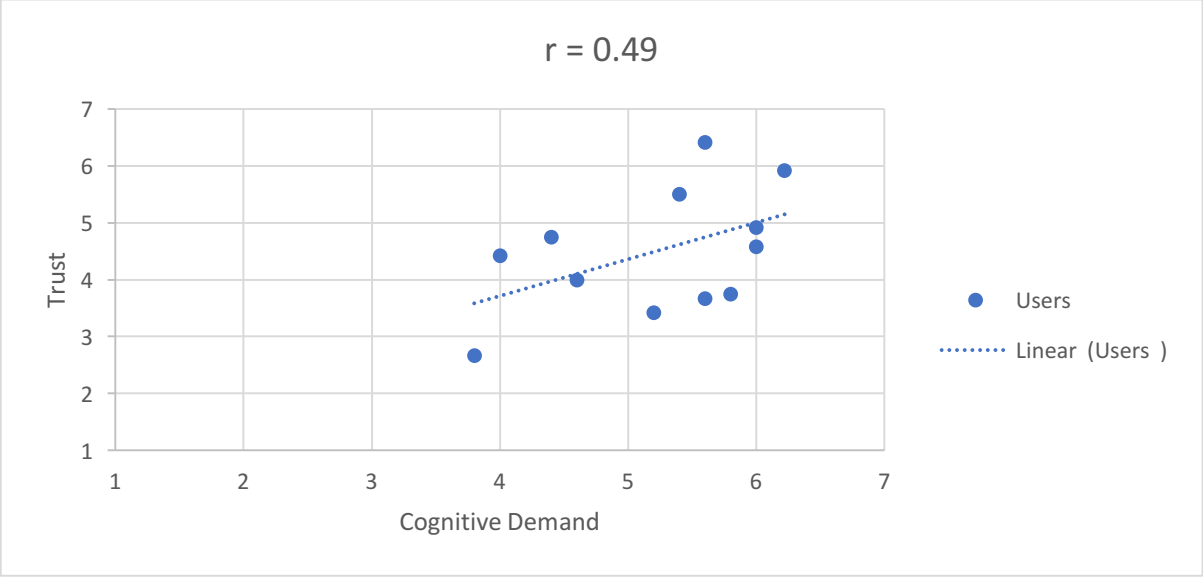
Likeability and Trust



Note: This graph shows the correlation of trust and likeability

Figure 20

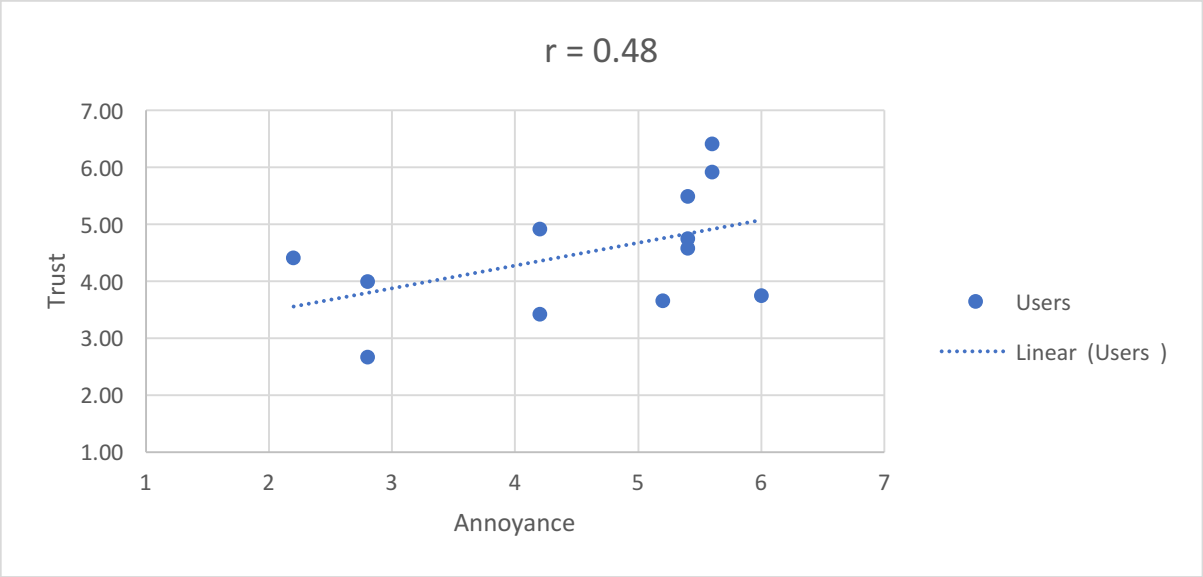
Cognitive Demand and Trust



Note: This graph shows the correlation of trust and cognitive demand

Figure 21

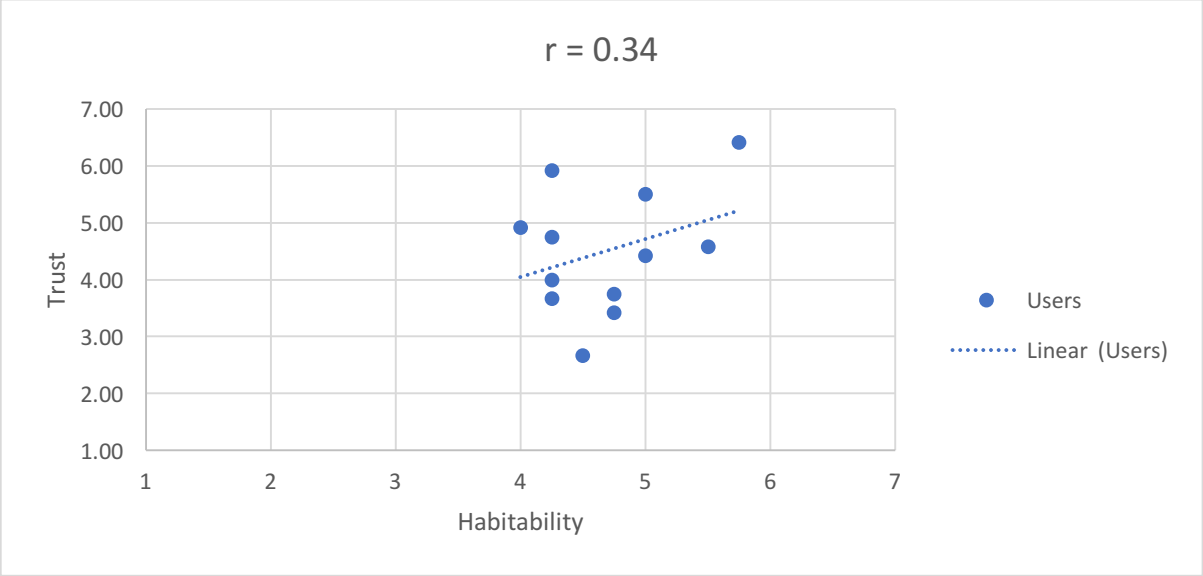
Annoyance and Trust



Note: This graph shows the correlation of trust and annoyance

Figure 22

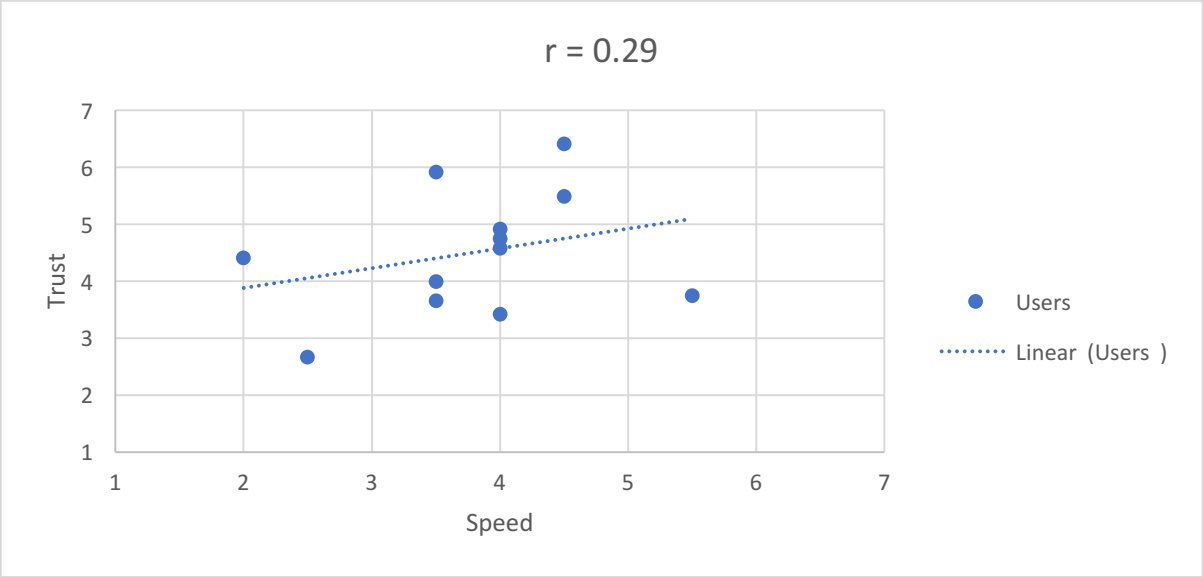
Habitability and Trust



Note: This graph shows the correlation of trust and habitability

Figure 23

Speed and Trust



Note: This graph shows the correlation of trust and speed

5. Discussion

This chapter will focus on discussing the results from the conducted research.

5.1. How do frequent users and infrequent users perceive voice assistants?

The literature review brought up the topics of trust (Clark et al., 2019), Luger & Sellen, 2016) and social presence/humanness of voice assistants and these were the topics (plus general usage) presented in the discussions, where certain themes emerged and showed similarities as well as differences between the two groups. From the initial discussions, it is evident that there are differences between frequent users and infrequent users however there does also seem to be similarities.

Both groups were asked questions on social presence and were asked to consider the humanlike characteristics of the voice assistant. The idea that the voice assistant could be humanlike was something that neither group had really considered with the infrequent users viewing it as more of a tool to help in accomplishing a goal and that changing a trait, such as the voice, was just like changing a setting on a computer. Only one participant (an infrequent user) had stated changing the assistant's voice in the past but they could not remember what they thought of the difference or why they did it. Questions on this topic were finalised with a discussion on their thoughts towards a more humanoid voice assistant. Frequent users enjoyed the idea and showed excitement towards it. They felt this would be a positive step closer to the world of science fiction whereas infrequent users found the idea invasive and felt that if the voice assistant could watch as well as listen, then they would be against it.

Both groups were then asked questions about trusting the voice assistant in regards to its capabilities as well as privacy. Frequent users showed a hesitation to use the voice assistant for anything other than carrying out simple tasks such as setting timers, playing a song or asking for the weather. Infrequent users had tried these simple tasks but felt the voice assistant could not carry out them to an acceptable standard. Frequent users mentioned failure with some tasks too, however, they acknowledged that the failure was due to how they interacted with the voice assistant and altered how they said their question, whereas with infrequent users, the inability to carry out the task initially was enough to avoid using it.

Infrequent users were reluctant to carry out any task involving personal information due to their discomfort of sharing too much with the voice assistant. Frequent users also showed some discomfort, however they had considered that the companies own a lot of data on them anyway, so they were not too worried. This fact (companies owning lots of data) was also brought up to the infrequent users to query why they still felt that way, but they could not figure out why they still felt uncomfortable.
Both sets of frequent users and infrequent users show similarities from the initial discussions. They both have issues with certain tasks that are able to be carried out by the voice assistant with both agreeing that they are only useful for simple tasks, however frequent users appear to have the knowledge of how to successfully interact with the voice assistant. The two groups also have concern for their privacy when around their voice assistant, however frequent users show acceptance of this due to the vast amounts of data that has already been collected on them after their usage of these companies (Google, Amazon, Apple etc.) products in the past.

These initial discussions indicate that frequent users will score higher in their perception of voice assistants than infrequent users and The Godspeed Questionnaire was utilised to explore these initial discussions further and quantify the difference between frequent users and infrequent users.
The results from the Godspeed Questionnaire (as seen in figure 1) show that frequent users showed a much higher score than infrequent users in the categories of animacy and likeability yet the differences (as seen in table 3). in anthropomorphism, perceived intelligence and perceived safety were small.
There was no real difference in regards to anthropomorphism ($p = 0.71$) and this could be due to the limitations of the product. They do not currently show physical human features and thus it is no surprise that both groups reported similar scores for anthropomorphism, however if this could have been changed then I believe there would have been an increase in the score as research has shown previously that anthropomorphising something increases trust in it (Yaseen & Lohan, 2018).
There was little difference in perceived safety ($p = 0.79$) which reflects the feedback from the focus groups/interviews. As previously mentioned, both groups showed consideration for the fact that the voice assistant could always be listening however frequent users were more accepting than infrequent users. This may be why there is no significant difference but a slightly higher score for frequent users.

The was also little difference in perceived intelligence ($p = 0.20$). The reason for this slight difference could be that both groups know where the information is coming from. Google, Amazon and Apple are all large companies with vast amounts of information. The device is not using its own stored information; it is collecting this data from a source that is well-known to both groups.

The largest differences were seen in likeability ($p < 0.0001$) and animacy ($p < 0.01$). In regards to animacy, much like anthropomorphism, the assistant is lacking features to make it feel alive such as movement, however, it does have a voice and arguably a personality. Frequent users and infrequent users both viewed the voice assistant as a tool, however it is possible that frequent users subconsciously viewed it as being "alive", despite not thinking this at the time of the discussions and is therefore the reason why there is a significant difference. In regards to likeability, the frequent users continue to use and have more experience with voice assistants whereas infrequent users have not used it as much and thus is likely why this difference is also significant.

Overall, the results show frequent users ($m = 3.58$) having a more positive perception of voice assistants than infrequent users ($m = 2.95$), however the difference ($p = 0.06$) between them is not significant enough to say that there is a definitive difference between the two groups.

5.2. How do frequent users and infrequent users trust voice assistants and how does this trust change through modifying the system's voice?

The results from the initial completion of The Checklist for Trust between People and Automation (as seen in table 7) show frequent users with an average trust score of 4.95 and infrequent users with an average trust score of 3.89 and the difference between the two can be considered significant ($p = 0.02$). The results from the second completion of The Checklist for Trust between People and Automation (as seen in table 7) show that frequent users had an average trust score of 5.20 whereas infrequent had an average trust score of 3.81 with a difference that can still be considered significant ($p = 0.02$). These results tell us, despite the two groups showing similarities in their perception of voice assistants, that there is a clear difference in their trust of voice assistants.

Frequent users saw their trust increase by 0.25 which cannot be considered a significant difference ($p = 0.32$) (as seen in table 8). Infrequent users saw their trust decrease by 0.08 which also cannot be considered a significant difference ($p = 0.86$) (as seen in table 8). As discussed previously, humanising a non-human product has proven to increase trust (Yaseen & Lohan, 2018), however, the results of my study have failed to reflect this in voice assistants. Frequent users did see a slight increase however the change was not significant. Infrequent users saw their trust decrease, but again, the change was not significant. However, something that was not taken into consideration was that other factors may have influenced the changes. Considering that a user's experience of using a technology has been shown to affect trust, as well as the user's own tendency to trust (Beldad et al., 2010) then changing the humanlike features of the voice assistant may not have been the only factor here. This could be why frequent users saw their trust increase at a higher rate than infrequent users, they were simply more experienced with the technology. Other factors like privacy/security and the reputation of the voice assistant's organisation could also have influenced the results: participants may have discovered information they did not know previously, or their opinions on certain companies could have changed.

Despite these group changes, individual results should also be examined. In regards to frequent users, there were increases for three participants whereas two participants had a decrease in trust, and one stayed the same. Regarding infrequent users, four participants reported a higher score with two participants scoring lower. The two infrequent users (P10 and P12) that lost trust in the voice assistant showed a much larger loss than the frequent users (P3 and P2) that lost trust. P10 and P12 scored an average of 1.37 less than their initial trust scores, whereas P3 and P2 scored an average of 0.29 less than their initial trust scores. This is most likely why the average infrequent trust score decreased despite more infrequent users showing a higher gain in trust. This suggests that when infrequent users lose trust they lose it more than frequent users do.

On the other hand, amongst those that did not lose trust, it is shown that more infrequent users gained an increase in trust (P7, P8, P9 and P11) (as seen in table 9) compared to frequent users (P1, P5 and P6) (as seen in table 10), with infrequent users increasing their trust by an average of 0.56 compared to frequent users increasing their trust by an average of 0.52. This suggests that when infrequent users gain trust, they gain it more than frequent users do.

Comparisons could also be made with the voice assistant that is being used and levels of trust. Users of Alexa showed the highest levels of trust yet they also showed the lowest levels of trust, the one Siri user showed the fourth highest trust level whereas Google Assistant was spread throughout the middle with only two participants scoring it less than half on the questionnaire. Of the four participants that saw a lower trust score, three had used Google Assistant, as well as the participant that reported no change. Only three (out of seven) users of Google Assistant users saw an increase, whereas three (out of four) Alexa users showed an increase. The single Siri user also saw an increase in their trust. However, despite these comparisons, they are not reliable. There was only one user of Siri, four users of Alexa and seven users of Google Assistant. These inconsistent numbers of different voice assistants being used make it hard to accept the comparisons that have been made and is something that I would reconsider if the research were to be repeated.

5.3. How does frequent users' and infrequent users' perception in voice assistants change through changing the system's voice?

Further observations can be made when using the results from second completion of the Godspeed Questionnaire. The results show (as seen in table 4) that, as a group, frequent users saw an increase of 0.08 whereas infrequent users, as a group, saw a decrease of 0.03. These changes are not significant enough (frequent: $p = 0.36$; infrequent: $p = 0.87$) to say that personalising the voice assistant's voice would also change how users perceive the system; however, these changes are similar to how participants' trust changed, with frequent users scoring higher and infrequent users scoring lower. Frequent users only saw animacy score slightly slower whereas infrequent users saw a slight decrease in anthropomorphism and a much larger decrease in perceived safety suggesting that even with a preferred voice, some still felt they were in danger. As stated before, it is very possible that other factors affected the changes in users. It is unlikely that users would show more concern for their safety with a voice they feel more comfortable with, so it is possible that these users who felt less safe based this from their thoughts around the security/privacy issues that are present with this technology.

One mistake I made was not tracking the specific user with their initial Godspeed results, and thus not being able to compare the perceptions from the before-experiment phase and the after-experiment phase. Had this been available, it would be possible to know if the same participants who saw a loss in trust also had a worse perception. Fortunately, I tracked participants' individual scores from the second completion of the Godspeed Questionnaire and these can be compared with the user's own trust scores to explore possible relations. A positive correlation is present between trust and perception amongst all users ($r = 0.69$) (as seen in figure 13) with frequent users showing a stronger correlation ($r = 0.68$) (as seen in figure 14) than infrequent users ($r = 0.47$) (as seen in figure 15). These results show that as a user's perception of a voice assistant increases, so does their trust in the system, which does suggest that the participants who saw their trust decrease may have been the participants that saw a decrease in perception.

5.4. To what extent does trust correlate with speech-system usability?

As in Beldad's review (Beldad et al., 2010), it is stated that if a product is usable then users will trust it more. My research aligns with this and these results can be seen in figure 17, 18, 19, 20, 21, 22 and 23.

Results of the SASSI represent accuracy, likeability, cognitive demand, annoyance, habitability and speed. These are compared with trust to understand if there is any correlation between trust and usability of voice assistants. The results show system response accuracy ($r = 0.66$), likeability ($r = 0.61$), cognitive demand ($r = 0.49$), annoyance ($r = 0.48$), habitability ($r = 0.34$), speed ($r = 0.29$) and overall usability ($r = 0.62$).

These results show that overall, there is a positive correlation between usability and trust (as seen in figure 17) and that each factor can be considered to positively correlate with trust, however some are weaker than others.

When evaluating voice assistant usability, several studies have focused on the quality of the system (Alagha & Helbing, 2019), (Berdasco et al., 2019), Kocabalil et al., 2018); however, there has not been much research into how this could affect trust. As mentioned, users show more trust towards a system if it returns correct and high quality content, but this has been shown through websites (Sillence et al., 2007). Aligning with this, the results (as seen in figure 18) show that there is indeed a positive correlation with user trust and system response accuracy of voice assistants. Like with accuracy, there is a moderate correlation ($r = 0.61$) with likeability (as seen in figure 19) and trust suggesting that if a user finds a system friendly and has an enjoyable experience in using the system then they will likely trust it.

A positive relationship with cognitive demand (as seen in figure 20) suggests if less effort is required from the user, then they will trust it more, however the weak correlation ($r = 0.49$) suggests that it doesn't have a strong effect on the user's trust.

A positive relationship with annoyance (as seen in figure 21) suggests that if the user is less frustrated with the voice assistant then they will trust it more, however again the weak correlation ($r = 0.48$) suggests that the impact it has on a user's trust is not large. A positive correlation with habitability (as seen in figure 22) suggests that if the user knows how to use the system, then they will show higher trust towards it, but like with cognitive demand and annoyance, the weak correlation ($r = 0.34$) suggests that there is no real link between habitability and trust. The weakest factor from the

questionnaire is speed, which shows that there is a very weak/no relationship (r = 0.29) between how quickly a voice assistant responds/interacts and how trustworthy the system is (as seen in figure 23).

The results suggest that to ensure higher trust in a voice assistant, designers must make sure that their system is enjoyable to use as well as providing the user with accurate responses. While the other factors of cognitive demand, annoyance, habitability and speed show a positive relationship with trust, they can all be considered weak in how they correlate with trust.

6. Conclusion

6.1. Project Summary

This study was conducted to understand how users perceive and trust voice assistants. To do this, participants were gathered into groups of frequent users and infrequent users to explore the differences that exists between the two in terms of how they use the technology as well as their thoughts. The participants were then instructed to use their voice assistant but with a preferable voice to understand if a more pleasant sounding voice assistant increased the user's trust. Their trust was measured before and after to compare the difference. Finally, the relationship that exists in perception and trust, as well as in usability and trust were explored too.

6.2. Aims

The main aim of this dissertation was to explore how users trust and perceive voice assistant technology, and I believe I have explored each aim well. They are summarised as follows:

One aim of this study was to explore the possible factors that could alter a user's interaction with voice assistant technology, and this was completed through a literature review.

The next aim was to discover how different frequent users and infrequent users are in how they use and perceive voice assistants and how this perception changes. This was explored through interviews and focus groups which brought up specific themes and ideas that were consistent amongst participants in their respective groups. These thoughts were then quantified through the Godspeed Questionnaire where results showed slight differences between the two groups. As expected, frequent users had a more positive perception of voice assistants than infrequent users did however the difference was not large enough to be considered different. I expected there to be a much larger difference between the two sets of users considering frequent users showed much more positivity towards voice assistants. Users took the questionnaire again after the experiment that showed slight changes in perception of both groups.

The next aim was to understand the development of a user's trust when personalising their voice assistant to have a preferable voice and this was explored through an experiment where users changed the voice assistant's default voice to a preferable one. Users initially completed The Checklist for Trust between People and Automation

and then after completed the same questionnaire. As groups, it showed that frequent users gained trust whereas infrequent users lost trust. However, as individuals, four infrequent users saw their trust increase whereas only three frequent users saw an increase, with two participants from each group losing trust. Infrequent users saw a more drastic loss in trust than frequent users, and this could explain why the infrequent group lost trust despite having more individuals gain it. Using trust and perception scores, a positive correlation between the two was revealed.

The final aim was to understand the relationships with trust and usability. Participants completed the SASSI and results were compared with their trust scores from the second completion of the Checklist for Trust between People and Automation. The results showed that as a user found a system more usable their trust increased. The strongest relationships of this were seen in system response accuracy and likeability. Results showed cognitive demand, annoyance, habitability and speed to all have weak correlations with trust.

6.3. Contributions

I believe I have contributed to an under researched topic (Edwards & Sanoubari, 2019), which made discussing my results against other papers in the domain of voice assistants slightly difficult. As stated before, trust has been explored in a number of industries such as in ecommerce or automation; however, there has not been much research into trust in the domain of voice assistants (Edwards & Sanoubari, 2019). There is also a lack of research comparing frequent users and infrequent users and the experiences each group has. Papers have looked at infrequent users (Cowan et al., 2017), and frequent users (Luger & Sellen, 2016), although much of this research is purely qualitative. Papers have also explored the quality of voice assistants (Alagha & Helbing, 2019), (Berdasco et al., 2019); however, to my knowledge, there are not any that explore trust as well. My research has also explored the relationship that exists between trust and speech-system usability of voice assistants. There has also been a lack of research exploring the changes when humanising a voice assistant. This has been present in robotics (Pollmann et al., 2020), (Yaseen & Lohan, 2018), yet it has not been fully explored in the domain of voice assistants. To my knowledge also, there has not been any studies exploring trust in voice assistants when changing the humanlike

characteristics of the device: there have been studies exploring user enjoyment (Wang et al., 2019), (Reinhardt et al., 2020), but not trust. I believe I have made several contributions to an area of voice assistant research that has yet to be properly explored.

6.4. Limitations

Due to the current global pandemic affecting the world, restrictions were placed upon us and much research was unable to be carried out. Facilities (such as universities) were closed, social distancing was in place, and being instructed by the government to only leave the home when necessary meant possible studies were impossible to be carried out. Because of this, the study had to be conducted online through channels like Skype and email, which meant I had to omit certain research aims from my study. One thing I wanted to achieve was to explore the three main voice assistants of Google Assistant, Alexa and Siri and understand which is more trustworthy. If I was to do this again, I would conduct it in a controlled environment such as the Sensorium at Edinburgh Napier University where I would be able to provide participants with the three voice assistants. One issue I faced, in regards to this, was finding anyone that owned all three so this exploration of trust in the three main voice assistants was not possible to conduct.

Another limitation could be the duration of the study. Users saw no significant changes to their trust after taking part in the study for a week, however, if more time was available then perhaps participants could have recorded their trust level every week and tracked trust on a weekly basis. Both sets of users may then see changes to their trust/perception until the point of where the scores start to stagnate.

Another limitation was sample size, and as stated before I would have liked to have explored trust in the different voice assistants. Ideally, I would have recruited more participants with more diverse voice assistant choice. Out of the twelve participants, seven were Google Assistant users, four were Alexa users and there was only one Siri user. If these numbers were more consistent, I could have had a deeper comparison into the trust between each device; however, the twelve participants in my study were the only ones to agree to take part and no one owned two different types of voice assistant.

I also believe that I could have managed my time better. I spent too long trying to figure out what my research would focus on, which put me behind by a week for most of the process. I did not consider submission dates for the research proposal or the initial report, and therefore did not include them on the timeline, had I done this then perhaps I would not have fell behind. The project timeline (as seen in Appendix D, showed three weeks to conduct my study; however, I would have added an extra week on to the experiment. As stated previously, users may have not seen significant changes to their trust over the week of using a personalised voice assistant, and one reason for this is the study was not long enough, and that more change could have been seen as users became more experienced and used to the new voice.

If carrying this research out again, I would assign more time to users interacting with this personalised voice assistant, as well as focusing on the type of voice assistant used for further observations.

6.5. Future Work

My study tried to understand if altering a humanlike characteristic (voice) in a non-human product could it appear more trustworthy. This experiment could be carried out again but could combine the personalised voice with an alternative body for the voice assistant, which is something that would be possible without social distancing. Trust could be measured amongst users using a voice assistant in its current form and then it could be measured again by hiding the voice assistant, but presenting users with a humanoid robot pretending that this is the voice assistant's form. Trust could then be compared with the non-humanoid appearance and the humanoid appearance. An animated humanoid appearance could also be utilised too to explore the inclusion of social cues, a feature that led users to prefer a humanoid agent over a non-humanoid agent (Reinhardt et al., 2020). As seen in a study comparing Pepper the robot with a voice assistant (Pollmann et al., 2020), in regards to pragmatic quality and hedonic quality, results showed that users reacted differently to the voice assistant, the non-animated Pepper, and the animated Pepper. Another study with Pepper also suggests that this could be useful where they measured trust between a humanoid robot and a machinelike robot (Yaseen & Lohan, 2018).

Several papers have explored the usability of different voice assistants, however there has not been any research into comparing trust in the different voice assistants, and

with trust known to be an issue in voice assistants, it is something that should be measured (Alagha & Helbing, 2019), (Berdasco et al., 2019).

6.6. Conclusion

Over duration of this dissertation, I have focused my research around voice assistants and how users trust and perceive this technology. I have researched relevant papers by conducting a literature review and have used those papers to shape my own piece of research where I carried out multiple processes of data gathering. I have examined established questionnaires and selected the appropriate tools to collect data for the area I have been exploring. I have organised and ran a small number of focus groups (a tool that I have never used before) and successfully conducted discussions with several participants. Upon collection of all this data, I analysed results and provided an interpretation on these results to provide research into an under-explored domain.

7. References

Alagha, E., & Helbing, R. (2019). Evaluating the quality of voice assistants' responses to consumer health questions about vaccines: an exploratory comparison of Alexa, Google Assistant and Siri. *BMJ Health & Care Informatics*, *26*(1), e100075. https://doi.org/10.1136/bmjhci-2019-100075

Anyoha, R. (2017). The History of Artificial Intelligence - Science in the News. Science in the News. Retrieved 2 June 2020, from http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence

Aylett, M., Cowan, B., & Clark, L. (2019). Siri, Echo and Performance. Extended Abstracts Of The 2019 CHI Conference On Human Factors In Computing Systems. https://doi.org/10.1145/3290607.3310422

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2008). Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal Of Social Robotics*, *1*(1), 71-81. https://doi.org/10.1007/s12369-008-0001-3

Beldad, A., de Jong, M., & Steehouder, M. (2010). How shall I trust the faceless and the intangible? A literature review on the antecedents of online trust. Computers In Human Behavior, 26(5), 857-869. https://doi.org/10.1016/j.chb.2010.03.013

Berdasco, López, Diaz, Quesada, & Guerrero. (2019). User Experience Comparison of Intelligent Personal Assistants: Alexa, Google Assistant, Siri and Cortana. Proceedings, 31(1), 51. https://doi.org/10.3390/proceedings2019031051

Bernhaupt, R., Murko, C., Pottier, G., & Battut, A. (2019). User Acceptance of Emotion-Aware Mood-Improving Voice Assistants. Retrieved 5 August 2020, from https://www.ibc.org/download?ac=10491.

Brooks, K., & Heuser, E. (2020). How one hospital is incorporating voice assistants into care delivery. Advisory.com. Retrieved 28 August 2020, from https://www.advisory.com/research/market-innovation-center/the-growth-channel/2020/03/voice-assistant-technology.

Caulfield, J. (2019). How to Do Thematic Analysis | A Step-by-Step Guide & Examples. Scribbr. Retrieved 7 August 2020, from https://www.scribbr.com/methodology/thematic-analysis/.

Clark, L., Munteanu, C., Wade, V., Cowan, B., Pantidi, N., & Cooney, O. et al. (2019). What Makes a Good Conversation?. *Proceedings Of The 2019 CHI Conference On Human Factors In Computing Systems - CHI '19*. https://doi.org/10.1145/3290605.3300705

Cowan, B., Pantidi, N., Coyle, D., Morrissey, K., Clarke, P., & Al-Shehri, S. et al. (2017). "What Can I Help You With?". Proceedings of The 19Th International Conference On Human-Computer Interaction With Mobile Devices And Services. https://doi.org/10.1145/3098279.3098539

Creswell, J. (2009). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (3rd ed., p. 211). SAGE Publications, Inc.

Cross-sectional vs. longitudinal studies. Iwh.on.ca. (2015). Retrieved 20 July 2020, from https://www.iwh.on.ca/what-researchers-mean-by/cross-sectional-vs-longitudinal-studies.

Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. Future Healthcare Journal, 6(2), 94-98. https://doi.org/10.7861/futurehosp.6-2-94

Druga, S., Williams, R., Breazeal, C., & Resnick, M. (2017). "Hey Google is it OK if I eat you?". Proceedings Of The 2017 Conference On Interaction Design And Children. https://doi.org/10.1145/3078072.3084330

Edwards, J., & Sanoubari, E. (2019). A need for trust in conversational interface research. *Proceedings Of The 1St International Conference On Conversational User Interfaces - CUI '19.* https://doi.org/10.1145/3342775.3342809

Egger, F. (2000). "Trust me, I'm an online vendor". CHI '00 Extended Abstracts On Human Factors In Computing Systems - CHI '00. https://doi.org/10.1145/633292.633352

Finnegan, M. (2020). 2020: The year the office finds its voice?. Computerworld. Retrieved 28 August 2020, from https://www.computerworld.com/article/3509470/2020-the-year-the-office-finds-its-voice.html.

Habler, F., Schwind, V., & Henze, N. (2019). Effects of Smart Virtual Assistants' Gender and Language. Proceedings Of Mensch Und Computer 2019 On - Muc'19. https://doi.org/10.1145/3340764.3344441

Hone, K., & Graham, R. (2000). Towards a tool for the Subjective Assessment of Speech System Interfaces (SASSI). Natural Language Engineering, 6(3&4), 287-303. https://doi.org/10.1017/s1351324900002497

Jian, J., Bisantz, A., & Drury, C. (2000). Foundations for an Empirically Determined Scale of Trust in Automated Systems. International Journal Of Cognitive Ergonomics, 4(1), 53-71. https://doi.org/10.1207/s15327566ijce0401_04

Kim, K., Boelling, L., Haesler, S., Bailenson, J., Bruder, G., & Welch, G. (2018). Does a Digital Assistant Need a Body? The Influence of Visual Embodiment and Social Behavior on the Perception of Intelligent Virtual Agents in AR. *2018 IEEE International Symposium On Mixed And Augmented Reality (ISMAR)*. https://doi.org/10.1109/ismar.2018.00039

Kiseleva, J., Williams, K., Jiang, J., Hassan Awadallah, A., Crook, A., Zitouni, I., & Anastasakos, T. (2016). Understanding User Satisfaction with Intelligent Assistants. Proceedings Of The 2016 ACM On Conference On Human Information Interaction And Retrieval - CHIIR '16. https://doi.org/10.1145/2854946.2854961

Kiseleva, J., Williams, K., Hassan Awadallah, A., Crook, A., Zitouni, I., & Anastasakos, T. (2016). Predicting User Satisfaction with Intelligent Assistants. Retrieved 10 August 2020, from https://www.microsoft.com/en-us/research/wp-content/uploads/2017/05/kiseleva_chiir2016_intelligent_satisfaction.pdf.

Kocabalil, A., Laranjo, L., & Coiera, E. (2018). Measuring User Experience in Conversational Interfaces: A Comparison of Six Questionnaires. https://doi.org/10.14236/ewic/hci2018.21

Lahoual, D., & Frejus, M. (2019). When Users Assist the Voice Assistants. Extended Abstracts Of The 2019 CHI Conference On Human Factors In Computing Systems. https://doi.org/10.1145/3290607.3299053

Lau, J., Zimmerman, B., & Schaub, F. (2018). Alexa, Are You Listening?. Proceedings Of The ACM On Human-Computer Interaction, 2(CSCW), 1-31. https://doi.org/10.1145/3274371

Lee, J., & See, K. (2004). Trust in Automation: Designing for Appropriate Reliance. Human Factors: The Journal Of The Human Factors And Ergonomics Society, 46(1), 50-80. https://doi.org/10.1518/hfes.46.1.50_30392

Lewis, J. (2016). Standardized Questionnaires for Voice Interaction Design. Association For Voice Interaction Design, 1(1). Retrieved 6 August 2020, from http://acixd.org/wp-content/uploads/2018/10/Standardized-Questionnaires-for-Voice-Interaction-Design.pdf.

Lovato, S., & Piper, A. (2015). "Siri, is this you?". Proceedings Of The 14Th International Conference On Interaction Design And Children - IDC '15. https://doi.org/10.1145/2771839.2771910

Luger, E., & Sellen, A. (2016). "Like Having a Really Bad PA". *Proceedings Of The 2016 CHI Conference On Human Factors In Computing Systems*. https://doi.org/10.1145/2858036.2858288

Malkin, N., Deatrick, J., Tong, A., Wijesekera, P., Egelman, S., & Wagner, D. (2019). Privacy Attitudes of Smart Speaker Users. Sciendo. Retrieved 10 June 2020, from https://petsymposium.org/2019/files/papers/issue4/popets-2019-0068.pdf.

Mayer, R., Davis, J., & Schoorman, F. (1995). An Integrative Model of Organizational Trust. The Academy Of Management Review, 20(3), 709. https://doi.org/10.2307/258792

Microsoft. (2019). Voice Report From answers to action: customer adoption of voice technology and digital assistants. Retrieved from https://about.ads.microsoft.com/en-us/insights/2019-voice-report

Mixed methods research. Resourcecentre.foodrisc.org. (2020). Retrieved 28 August 2020, from http://resourcecentre.foodrisc.org/mixed-methods-research_185.html.

McGill University. (2018). Do we trust people who speak with an accent? We tend to believe speakers who sound the same as us, though much depends on their tone of voice. ScienceDaily. Retrieved July 24, 2020 from www.sciencedaily.com/releases/2018/09/180918082049.htm

Myers, C., Grethlein, D., Furqan, A., Ontañón, S., & Zhu, J. (2019). Modeling Behavior Patterns with an Unfamiliar Voice User Interface. *Proceedings Of The 27Th ACM Conference On User Modeling, Adaptation And Personalization*. https://doi.org/10.1145/3320435.3320475

Norman, D., & Nielsen, J. The Definition of User Experience (UX). Nielsen Norman Group. Retrieved 16 July 2020, from https://www.nngroup.com/articles/definition-user-experience/.

Orehovački, T., Babić, S., & Etinger, D. (2017). Modelling the Perceived Pragmatic and Hedonic Quality of Intelligent Personal Assistants. Intelligent Human Systems Integration, 589-594. https://doi.org/10.1007/978-3-319-73888-8_91

Plummer, L. Are AI Voice Assistants the Next Big Edtech Trend?. Intel. Retrieved 28 August 2020, from https://www.intel.co.uk/content/www/uk/en/it-managers/education-voice-technology.html.

Porcheron, M., Fischer, J., Reeves, S., & Sharples, S. (2018). Voice Interfaces in Everyday Life. *Proceedings Of The 2018 CHI Conference On Human Factors In Computing Systems - CHI '18*. https://doi.org/10.1145/3173574.3174214

Pollmann, K., Ruff, C., Vetter, K., & Zimmermann, G. (2020). Robot vs. Voice Assistant. Companion Of The 2020 ACM/IEEE International Conference On Human-Robot Interaction. https://doi.org/10.1145/3371382.3378251

Rajkumar, R. (2015). Self-Driving Vehicles. *Proceedings Of The 13Th ACM Conference On Embedded Networked Sensor Systems - Sensys '15*. https://doi.org/10.1145/2809695.282346

Reeves, S., Porcheron, M., Fischer, J., Candello, H., McMillan, D., & McGregor, M. et al. (2018). Voice-based Conversational UX Studies and Design. *Extended Abstracts Of The 2018 CHI Conference On Human Factors In Computing Systems*. https://doi.org/10.1145/3170427.3170619

Reinhardt, J., Hillen, L., & Wolf, K. (2020). Embedding Conversational Agents into AR. Proceedings Of The Fourteenth International Conference On Tangible, Embedded, And Embodied Interaction. https://doi.org/10.1145/3374920.3374956

Ren, X. (2016). Rethinking the Relationship between Humans and Computers. *Computer*, *49*(8), 104-108. https://doi.org/10.1109/mc.2016.253

Sano, S., Kaji, N., & Sassano, M. (2016). Prediction of Prospective User Engagement with Intelligent Assistants. Proceedings Of The 54Th Annual Meeting Of The

Association For Computational Linguistics (Volume 1: Long Papers). https://doi.org/10.18653/v1/p16-1114

Sauro, J. (2015). *5 Types of Qualitative Methods*. Measuringu.com. Retrieved 20 July 2020, from https://measuringu.com/qual-methods/.

Sayago, S., Neves, B., & Cowan, B. (2019). Voice assistants and older people. Proceedings Of The 1St International Conference On Conversational User Interfaces - CUI '19. https://doi.org/10.1145/3342775.3342803

Sciuto, A., Saini, A., Forlizzi, J., & Hong, J. (2018). "Hey Alexa, What's Up?". Proceedings Of The 2018 On Designing Interactive Systems Conference 2018 - DIS '18. https://doi.org/10.1145/3196709.3196772

Shi, Y., Yan, X., Ma, X., Lou, Y., & Cao, N. (2018). Designing Emotional Expressions of Conversational States for Voice Assistants. Extended Abstracts Of The 2018 CHI Conference On Human Factors In Computing Systems. https://doi.org/10.1145/3170427.3188560

Shorten, A., & Smith, J. (2017). Mixed methods research: expanding the evidence base. *Evidence Based Nursing*, *20*(3), 74-75. https://doi.org/10.1136/eb-2017-102699

Siegal, J. (2019). *Google admits over 1,000 private Google Home recordings were leaked*. New York Post. Retrieved 29 June 2020, from https://nypost.com/2019/07/12/google-admits-over-1000-private-google-home-recordings-were-leaked/.

Sillence, E., Briggs, P., Harris, P., & Fishwick, L. (2007). Health Websites that people can trust – the case of hypertension. Interacting With Computers, 19(1), 32-42. https://doi.org/10.1016/j.intcom.2006.07.009

Smith, A. (2018). Conversational UX Design : An Introduction - Usability Geek. Usability Geek. Retrieved 6 June 2020, from https://usabilitygeek.com/conversational-ux-design-introduction/.

Statista Research Department (2019). Number of voice assistants in use worldwide 2019-2023 | Statista. Statista. (2019). Retrieved 2 June 2020, from https://www.statista.com/statistics/973815/worldwide-digital-voice-assistant-in-use/.

Strengers, Y., Kennedy, J., Arcari, P., Nicholls, L., & Gregg, M. (2019). Protection, Productivity and Pleasure in the Smart Home. *Proceedings Of The 2019 CHI Conference On Human Factors In Computing Systems - CHI '19*. https://doi.org/10.1145/3290605.3300875

Sundar, S., Jung, E., Waddell, T., & Kim, K. (2017). Cheery companions or serious assistants? Role and demeanor congruity as predictors of robot attraction and use intentions among senior citizens. *International Journal Of Human-Computer Studies*, *97*, 88-97. https://doi.org/10.1016/j.ijhcs.2016.08.006

Triantafyllidou, D., & Tefas, A. (2016). Face detection based on deep convolutional neural networks exploiting incremental facial part learning. *2016 23Rd International Conference On Pattern Recognition (ICPR)*. https://doi.org/10.1109/icpr.2016.7900186

Turing, A., 1950. I.—Computing Machinery and Intelligence. Mind, LIX(236), pp.433-460. https://doi.org/10.1093/mind/lix.236.433/

Völkel, S., Kempf, P., & Hussmann, H. (2020). Personalised Chats with Voice Assistants. Proceedings Of The 2Nd Conference On Conversational User Interfaces. https://doi.org/10.1145/3405755.3406156

Wang, I., Smith, J., & Ruiz, J. (2019). Exploring Virtual Agents for Augmented Reality. Proceedings Of The 2019 CHI Conference On Human Factors In Computing Systems - CHI '19. https://doi.org/10.1145/3290605.3300511

Where is Artificial Intelligence Used Today?. (2020). Retrieved 2 March 2020, from https://becominghuman.ai/where-is-artificial-intelligence-used-today-3fd076d15b68

Yaseen, A., & Lohan, K. (2018). Playing Pairs with Pepper. Retrieved 10 August 2020, from https://arxiv.org/pdf/1810.07593.pdf.

Zhou, T. (2013). Understanding user adoption of location-based services from a dual perspective of enablers and inhibitors. Information Systems Frontiers, 17(2), 413-422. https://doi.org/10.1007/s10796-013-9413-1

8. Appendices

8.1. Appendix A: Godspeed Questionnaire

GODSPEED I: ANTHROPOMORPHISM

Please rate your impression of the robot on these scales:

Fake 1 2 3 4 5 Natural
Machinelike 1 2 3 4 5 Humanlike
Unconscious 1 2 3 4 5 Conscious
Artificial 1 2 3 4 5 Lifelike
Moving rigidly 1 2 3 4 5 Moving elegantly

GODSPEED II: ANIMACY

Please rate your impression of the robot on these scales:

Dead 1 2 3 4 5 Alive
Stagnant 1 2 3 4 5 Lively
Mechanical 1 2 3 4 5 Organic
Artificial 1 2 3 4 5 Lifelike
Inert 1 2 3 4 5 Interactive
Apathetic 1 2 3 4 5 Responsive

GODSPEED III: LIKEABILITY

Please rate your impression of the robot on these scales:

Dislike 1 2 3 4 5 Like
Unfriendly 1 2 3 4 5 Friendly
Unkind 1 2 3 4 5 Kind
Unpleasant 1 2 3 4 5 Pleasant
Awful 1 2 3 4 5 Nice

GODSPEED IV: PERCEIVED INTELLIGENCE

Incompetent 1 2 3 4 5 Competent
Ignorant 1 2 3 4 5 Knowledgeable
Irresponsible 1 2 3 4 5 Responsible
Unintelligent 1 2 3 4 5 Intelligent
Foolish 1 2 3 4 5 Sensible

GODSPEEC V: PERCEIVED SAFETY

Anxious 1 2 3 4 5 Relaxed
Agitated 1 2 3 4 5 Calm
Quiescent 1 2 3 4 5 Surprised

8.2. Appendix B: The Checklist for Trust between People and Automation

Checklist for Trust between People and Automation

Below is a list of statement for evaluating trust between people and automation. There are several scales for you to rate intensity of your feeling of trust, or your impression of the system while operating a machine. Please mark an "x" on each line at the point which best describes your feeling or your impression.

(Note: not at all=1; extremely=7)

1 The system is deceptive

1 2 3 4 5 6 7

2 The system behaves in an underhanded manner

1 2 3 4 5 6 7

3 I am suspicious of the system's intent, action, or outputs

1 2 3 4 5 6 7

4 I am wary of the system

1 2 3 4 5 6 7

5 The system's actions will have a harmful or injurious outcome

1 2 3 4 5 6 7

6 I am confident in the system

1 2 3 4 5 6 7

7 The system provides security

1 2 3 4 5 6 7

8 The system has integrity

1 2 3 4 5 6 7

9 The system is dependable

1 2 3 4 5 6 7

10 The system is reliable

1 2 3 4 5 6 7

11 I can trust the system

1 2 3 4 5 6 7

12 I am familiar with the system

1 2 3 4 5 6 7

8.3. Appendix C: Subjective Assessment for Speech System Interfaces

		Strongly disagree	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Strongly agree
System Response Accuracy	1. The system is accurate.	○	○	○	○	○	○	○
	2. The system is unreliable.	○	○	○	○	○	○	○
	3. The interaction with the system is unpredictable.	○	○	○	○	○	○	○
	4. The system didn't always do what I wanted.	○	○	○	○	○	○	○
	5. The system didn't always do what I expected.	○	○	○	○	○	○	○
	6. The system is dependable.	○	○	○	○	○	○	○
	7. The system makes few errors.	○	○	○	○	○	○	○
	8. The interaction with the system is consistent.	○	○	○	○	○	○	○
	9. The interaction with the system is efficient.	○	○	○	○	○	○	○
Likeability	10. The system is useful.	○	○	○	○	○	○	○
	11. The system is pleasant.	○	○	○	○	○	○	○
	12. The system is friendly.	○	○	○	○	○	○	○
	13. I was able to recover easily from errors.	○	○	○	○	○	○	○
	14. I enjoyed using the system.	○	○	○	○	○	○	○
	15. It is clear how to speak to the system.	○	○	○	○	○	○	○
	16. It is easy to learn to use the system.	○	○	○	○	○	○	○
	17. I would use this system.	○	○	○	○	○	○	○
	18. I felt in control of the interaction with the system.	○	○	○	○	○	○	○
Cognitive Demand	19. I felt confident using the system.	○	○	○	○	○	○	○
	20. I felt tense using the system.	○	○	○	○	○	○	○
	21. I felt calm using the system.	○	○	○	○	○	○	○
	22. A high level of concentration is required when using the system.	○	○	○	○	○	○	○
	23. The system is easy to use.	○	○	○	○	○	○	○
Annoyance	24. The interaction with the system is repetitive.	○	○	○	○	○	○	○
	25. The interaction with the system is boring.	○	○	○	○	○	○	○
	26. The interaction with the system is irritating.	○	○	○	○	○	○	○
	27. The interaction with the system is frustrating.	○	○	○	○	○	○	○
	28. The system is too inflexible.	○	○	○	○	○	○	○
Habitability	29. I sometimes wondered if I was using the right word.	○	○	○	○	○	○	○
	30. I always knew what to say to the system.	○	○	○	○	○	○	○
	31. I was not always sure what the system was doing.	○	○	○	○	○	○	○
	32. It is easy to lose track of where you are in an interaction with the system.	○	○	○	○	○	○	○
Speed	33. The interaction with the system is fast.	○	○	○	○	○	○	○
	34. The system responds too slowly.	○	○	○	○	○	○	○

8.4. Appendix D: Project Timeline

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7
Research proposal							
Literature review							

	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13	Week 14
Methodology							
Carry out study							
Analyse results							
Discuss results							
Finalise & submit							

8.5. Appendix E: Application for Cross-University Ethical Approval

Name:	Liam Rawsthorne
School or Professional service department:	School of Computing
Email:	40085161@napier.live.ac.uk
Contact number:	07562801717
Project Title:	How Do Users Perceive and Trust Voice Assistant Technology?
Start Date:	18/05/2020
Duration of Project:	14 weeks
Type of Research: PG/Masters/Student	

Screening Questions

Please answer the following questions to identify the level of risk in the proposed project:

If you answer 'No' to all questions, please complete Section 3a only.

If you have answered 'Yes' to any of the questions 5-14 please complete Section 3a and 3b.

If you have answered 'Yes to any of the questions 1-4, complete all of Section 3.

	You Must Answer All Questions	**Yes**	**No**
1.	Is the research clinical in nature?	☐	☒
2.	Is the research investigating socially or culturally 'controversial' topics (for example pornography, extremist politics, or illegal activities)?	☐	☒
3.	Will any covert research method be used?	☐	☒
4.	Will the research involve deliberately misleading participants (deception) in any way?	☐	☒
5.	Does the Research involve staff or students within the University?	☒	☐
6.	Does the Research involve vulnerable people? (For example people under 18 or over 70 years of age, disabled (either physically	☐	☒

	or mentally), those with learning difficulties, people in custody, migrants etc).		
7.	Is the information gathered from participants of a sensitive or personal nature?	☐	☒
8.	Is there any realistic risk of any participants experiencing either physical or psychological distress or discomfort?	☐	☒
9.	Have you identified any potential risks to the researcher in carrying out the research? (for example physical/emotional/social/economic risks?)	☐	☒
10.	Are there implications from a current or previous professional relationship i.e. staff/student/line manager/managerial position that would affect the voluntary nature of the participation?	☐	☒
11.	Will the research require the use of assumed consent rather than informed consent? (For example when it may be impossible to obtain informed consent due to the setting for the research – e.g. observational studies/videoing/photography within a public space)	☐	☒
12.	Is there any risk to respondents' anonymity in any report/thesis/publication from the research, even if real names are not used?	☐	☒
13.	Will any payment or reward be made to participants, beyond reimbursement or out-of-pocket expenses?	☐	☒
14.	Does the research require external ethics clearance? (For example from the NHS or another institution)	☐	☒
15.	Does the research involve the use of secondary data?	☐	☒

3A. Details of Project

In this section please provide details of your project and outline data collection methods, how participant consent will be given as well as details of storage and dissemination.

Please give a 300 word overview of the research project
With the rise in popularity of voice assistants such as Alexa and Google Assistant, it is important to consider the user experience surrounding these systems. This project aims to explore the

field of conversational artificial intelligence (CAI) and certain concepts such as trust and social presence and dimensions such as engagement, emotion and pragmatic quality to help understand what affects the conversational user experience (CUX) of using a voice interface.

The project will look to answer the following research questions:

- What are the factors influencing interaction between people and CAI?
- How do frequent users and infrequent users perceive voice assistants?
- How does frequent users' and infrequent users trust in voice assistants change through personalisation of voice?
- How does trust relate with usability?

A study will be carried out with participants consisting of frequent users and infrequent users to understand the differences between the people that do interact with the technology and those who choose not to. Frequent users/infrequent users will be interviewed before to find out their current thoughts and these will be compared. During the study, they will complete tasks and fill out questionnaires which will measure perception, trust and usability. This will assist in helping to understand the experience users have with current voice assistant systems.

Data Collection	
1.	**Who will be the participants in the research?**
	Frequent Users vs Infrequent Users of voice assistant technology
2.	**How will you collect and analyse the research data? (please outline all methods e.g. questionnaires/focus groups/internet searches/literature searches/interviews/observation)**
	Literature review, interviews, focus groups, questionnaires, thematic analysis, statistical analysis, descriptive statistics
3.	**Where will the data will be gathered (e.g. in the classroom/on the street/telephone/on-line)**
	Over video call, on-line, in the home

4.	**Please describe your selection criteria for inclusion of participants in the study**
	My study involves exploring the factors that affect the user experience of voice assistants. For this I must ask users for their experiences and thoughts surrounding voice assistants
5.	**If your research is based on secondary data, please outline the source, validity and reliability of the data set**
	N/A
Consent and Participant Information	
7.	**How will you invite research participants to take part in the study? (e.g. letter/email/asked in lecture)**
	Email
8.	**How will you explain the nature and purpose of the research to participants?**
	Through a participant information sheet sent via email
9.	**How will you record obtaining informed consent from your participants?**
	Consent forms will be sent via email and sent back with consent
Data storage and Dissemination	
10.	**How and in what format will data be stored? And what steps will be taken to ensure data is stored securely?**
	Files will be stored online in a protected google drive
11.	**Who will have access to the data?**
	Only I will have access

12.	**Will the data be anonymised so that files contain no information that could be linked to any participant?**
	Yes
13.	**How long will the data be kept?**
	Until no longer needed
14.	**What will be done with the data at the end of the project?**
	The data will be deleted
15.	**How will the findings be disseminated?**
	They will be explained in my dissertation study
16.	**Will any individual be identifiable in the findings?**
	No

3B. Identification and Mitigation of Potential risks

This section is designed to identify any realistic risks to the participants and how you propose to deal with it.

1. **Does this research project involve working with potentially vulnerable individuals?**

Group	**Yes**	**NO**	**Details (for example programme student enrolled on, or details of children's age/care situation, disability)**
Students at Napier	☒	☐	
Staff at ENU	☐	☒	
Children under 18	☐	☒	
Elderly (over 70)	☐	☒	
Disabled	☐	☒	

Migrant workers	☐	☒	
Prisoners / people in custody	☐	☒	
Learning difficulties	☐	☒	

2. If you are recruiting children (under 18 years) or people who are otherwise unable to give informed consent, please give full details of how you will obtain consent from parents, guardians, carers etc.

N/A

3. Please describe any identified risks to participants or the researcher as a result of this research being carried out

N/A

4. Please describe what steps have been taken to reduce these identified risks? (for example providing contact details for appropriate support services (e.g. University Counselling, Samaritans), reminding participants of their right to withdraw and/or not answering questions, or providing a full debriefing to participants)

N/A

5. If you plan to use assumed consent rather than informed consent please outline why this is necessary

N/A

6. **If payment or reward will be made to participants please justify that the amount and type are appropriate (for example the amount should not be so high that participants would be financially coerced into taking part, or that the type of reward is appropriate to the research topic).**

N/A

3C. Justification of High Risk Projects

If you answered 'Yes' to the screening questions 1-4 this section asks for justification on the choice of research topic and methodology.

1. **If you have answered yes to question 1 please give a full description of all medical procedures to be used within the research and provide evidence that the project has obtained NHS ethical approval.**

N/A

2. **If you have answered yes to questions 2 (research into a controversial topic) please provide a justification for your choice of research topic, and describe how you would deal with any potential issues arising from researching that topic.**

N/A

3. **If you have answered yes to questions 3 or 4 (use of deception or covert research methods) please provide a justification for your choice of methodology, and state how you will mitigate the risks associated with these approaches**.

N/A

Declaration	
☒	I consider that this project has no significant ethical implications to be brought to the attention of Research Integrity Committee
☐	I consider that this project may have significant ethical implications to be brought to the attention of the Research Integrity Committee
Researcher Signature: Liam Rawsthorne	**Date: 30/06/20**
Director of Studies/Supervisor/Principal Investigator Signature: Emilia Sobolewska	**Date: 30/06/20**

Checklist

All applications require the following to be submitted with the application form

Participant Information Sheet	☒
Informed Consent Form	☒
Interview/Survey Questions	☒

8.6. Appendix F: Interview/Focus Group Participant Information Sheet

Participant Information Project: How Do Users Perceive and Trust Voice Assistant Technology?

The aim of the study is to explore the factors that may affect the conversational user experience. This focus group/interview aims to collect qualitative data from users based on concepts surrounding voice assistants.

The session is not expected to last more than __ minutes. It will be held online and will require a computer/laptop with a microphone and working internet connection. The discussion will be recorded.

Upon the publication of data, all participants will be anonymised and names will not be linked with the research materials. Data will be securely stored until the project finishes and upon completion will be permanently destroyed.

Participation is voluntary and participants may withdraw at any point

If there are any questions regarding the study, please contact me through 40085161@live.napier.ac.uk

8.7. Appendix G: Interview/Focus Group Participation Consent

Project: How Do Users Perceive and Trust Voice Assistant Technology?

Edinburgh Napier University requires that all persons who participate in research studies give their written consent to do so. Please read the following and sign it if you agree with what it says.

1. I freely and voluntarily consent to be a participant in this research to be conducted by Liam Rawsthorne, who is a postgraduate student in the Edinburgh Napier School of Computing.

2. I have been informed of the broad goal of this research study. I have been told what is expected of me and that the study should take no longer than 30 minutes to complete.

3. I have been told that my responses will be anonymised. My name will not be linked with the research materials, and I will not be identified or identifiable in any report subsequently produced by the researcher. I have been told that these data may be submitted for publication.

4. I also understand that if at any time during the session:
   If I feel unable or unwilling to continue, I am free to leave. That is, my participation in this study is completely voluntary, and I may withdraw from it at any time without negative consequences.

5. In addition, should I not wish to answer any particular question or questions, I am free to decline.

6. I have been given the opportunity to ask questions regarding the session and my questions have been answered to my satisfaction.

7. I have read and understand the above and consent to participate in this study. My signature is not a waiver of any legal rights. Furthermore, I understand that I will be able to keep a copy of this consent form for my records.

Participant's Signature :
Date :

I have explained and defined in detail the research procedure in which the respondent has consented to participate. Furthermore, I will retain one copy of the informed consent form for my records.

Researcher's Signature:
Date:

8.8. Appendix H: Online Forms Consent - Trust Questionnaire (1)

Checklist for Trust between People and Automation

I am conducting a study that explores the user experience of voice assistants (Siri, Google Assistant, Alexa etc) and the factors that may affect a user's interaction with the technology.

This questionnaire aims to explore the human-machine trust regarding voice assistants. This will be taken again later on to compare results and to assist in understanding.

If any questions are required, please contact me at the following:
40085161@live.napier.ac.uk

* Required

First Name (For comparison) *

Your answer

Consent Statement

Please read the following and if you agree fill in the checkbox below

I freely and voluntarily consent to be a participant in this research to be conducted by Liam Rawsthorne, who is a postgraduate student at the Edinburgh Napier School of Computing.

My responses will be anonymised. My name will not be linked with the research materials, and I will not be identified or identifiable in any report subsequently produced by the researcher. I understand that this data may be submitted for publication.

I understand that if at any time during the survey, I feel unable or unwilling to continue, I am free to stop.

I have been given the opportunity to ask questions regarding the questionnaire and my questions have been answered to my satisfaction.

I understand that I have the right to withdraw consent and ask for my data to be removed.

Consent *

☐ I understand the above statement and give consent to participate in this questionnaire

8.9. Appendix I: Online Forms Consent - Trust Questionnaire (2)

Checklist for Trust between People and Automation

I am conducting a study that explores the user experience of voice assistants (Siri, Google Assistant, Alexa etc) and the factors that may affect a user's interaction with the technology.

This questionnaire aims to explore the human-machine trust regarding voice assistants.

If any questions are required, please contact me at the following:
40085161@live.napier.ac.uk

* Required

First Name (For comparison) *

Your answer

Consent Statement

Please read the following and if you agree fill in the checkbox below

I freely and voluntarily consent to be a participant in this research to be conducted by Liam Rawsthorne, who is a postgraduate student at the Edinburgh Napier School of Computing.

My responses will be anonymised. My name will not be linked with the research materials, and I will not be identified or identifiable in any report subsequently produced by the researcher. I understand that this data may be submitted for publication.

I understand that if at any time during the survey, I feel unable or unwilling to continue, I am free to stop.

I have been given the opportunity to ask questions regarding the questionnaire and my questions have been answered to my satisfaction.

I understand that I have the right to withdraw consent and ask for my data to be removed.

Consent *

☐ I understand the above statement and give consent to participate in this questionnaire

8.10. Appendix J: Online Forms Consent - Perception Questionnaire (1 and 2)

Godspeed Questionnaire

I am conducting a study that explores the user experience of voice assistants (Siri, Google Assistant, Alexa etc) and the factors that may affect a user's interaction with the technology.

This questionnaire aims to gather data on your current perception of voice assistants.

If any questions are required, please contact me at the following:
40085161@live.napier.ac.uk

* Required

Consent Statement

Please read the following and if you agree fill in the checkbox below

I freely and voluntarily consent to be a participant in this research to be conducted by Liam Rawsthorne, who is a postgraduate student at the Edinburgh Napier School of Computing.

My responses will be anonymised. My name will not be linked with the research materials, and I will not be identified or identifiable in any report subsequently produced by the researcher. I understand that this data may be submitted for publication.

I understand that if at any time during the survey, I feel unable or unwilling to continue, I am free to stop.

I have been given the opportunity to ask questions regarding the questionnaire and my questions have been answered to my satisfaction.

I understand that I have the right to withdraw consent and ask for my data to be removed.

Consent *

☐ I understand the above statement and give consent to participate in this questionnaire

8.11. Appendix K: Online Forms Consent - SASSI Questionnaire

Subjective Assessment of Speech System Interfaces (SASSI)

I am conducting a study that explores the user experience of voice assistants (Siri, Google Assistant, Alexa etc) and the factors that may affect a user's interaction with the technology.

This questionnaire aims to gather data on general speech-system usability of voice assistants

If any questions are required, please contact me at the following:
40085161@live.napier.ac.uk

* Required

First Name (For comparison) *

Your answer

Consent Statement

Please read the following and if you agree fill in the checkbox below

I freely and voluntarily consent to be a participant in this research to be conducted by Liam Rawsthorne, who is a postgraduate student at the Edinburgh Napier School of Computing.

My responses will be anonymised. My name will not be linked with the research materials, and I will not be identified or identifiable in any report subsequently produced by the researcher. I understand that this data may be submitted for publication.

I understand that if at any time during the survey, I feel unable or unwilling to continue, I am free to stop.

I have been given the opportunity to ask questions regarding the questionnaire and my questions have been answered to my satisfaction.

I understand that I have the right to withdraw consent and ask for my data to be removed.

Consent *

☐ I understand the above statement and give consent to participate in this questionnaire

8.12. Appendix L: Original Proposal

The process of completing and reviewing the contents of this form is intended ensure that the proposed project is viable. It is also intended to increase the chances of a good pass. Much of the material produced while completing this form may be reused in the dissertation itself.

1. **Student details**

First name	Liam
Last (family) name	Rawsthorne
Edinburgh Napier matriculation number	40085161

2. **Details of your programme of study**

MSc Programme title	MSc Computing (User Experience)
Year that you started your diploma modules	2019

3. **Project outline details**

Please suggest a title for your proposed project. If you have worked with a supervisor on this proposal, please provide the name. You are strongly advised to work with a member of staff when putting your proposal together.

Title of the proposed project	Evaluating the User Experience of Conversational Artificial Intelligence
Is your project appropriate to your programme of study?	Yes
Name of supervisor	Emilia Sobolewska

4. **Brief description of the research area - background**

Please do not describe your project in this section. Instead, provide background information in the box below on the broad research area in which your project sits. You should write in narrative (not bullet points). The academic/theoretical basis of your description of the research area should be evident through the use of citations and references. Your description should be between half and one page in length.

An example of artificial intelligence (AI) that is currently used by many and expected to grow is conversational artificial intelligence (CAI). CAI can be seen in the form of voice user interfaces (VUIs) in the home via digital assistants such as Siri, Google Assistant and Alexa, as well as existing in smart devices such as mobile phones, tablets and watches. In 2019 it was estimated that, worldwide, 3.25 billion devices used a voice assistant (e.g. Google Assistant, Siri, Cortana) and in 2023 this is expected to rise to around 8 billion (Statista Research Department, 2019).

Not only is this growth seen in digital assistants out of and within the home, but it can also be seen through the popularity of chatbots. In retail, increasing customer demand is an important reason for business to consider using a chatbot. More than half of customers expect a business to be available 24/7 and 65% would rather use a messaging app rather to get in touch with a business rather than over the phone (Chatbots 101, 2017). The success rate of chatbots in healthcare is

expected to rise to 75% in 2022 (from 12% in 2017), and in banking it is predicted to rise to 90% in the same year ("Chatbots, a Game Changer for Banking & Healthcare, Saving $8 billion Annually by 2022", 2017). It is evident that digital assistants and chatbots will play an integral part of our future lives.

Conversational User Experience (CUX) is the user experience (UX) of any natural language based technology that replicates a human conversation and good CUX should make the flow between two people feel natural, even if one is a VUI or a bot (Morphy, 2018). When designing a CUX, understanding how human conversation is structured is necessary, the CAI should be designed to make the experience as realistic as talking to a human and to be as natural as possible (Moore & Arar, 2019). One UX designer from the Amazon Alexa team stated that users found it strange because they were interacting with a voice that wasn't in the same room, and that the best way to make users forget the device is a robot is to make it more humanlike (Nguyen, 2018). If it doesn't feel natural, users may find it awkward to talk to the CAI and therefore will be less inclined to use the service (Smith, 2018). A more humanlike CAI may also help to improve a user's social presence, as seen when users reported a higher level of it when a digital assistant was given a human avatar to go with the voice (Kim et al., 2018).

The CUX is also affected by concepts of UX such as user engagement and trust. User engagement is concerned with helping the user to complete a goal (Sutcliffe, 2016), and if they are interacting with the CAI, they are likely using it to complete a task for convenience (Smith, 2018) and the CAI should be designed with this in mind. Practices such as focusing on conversational flow, acknowledging errors and making the experience personal (Lean, 2019) can be employed to improve the CUX.

To interact with the CAI, the user must give over personal data and users must be able to trust whatever is collecting it. Users have demonstrated concern with what happens to their data in terms of ownership (Cowan et al., 2017), and how it is stored (Malkin et al., 2019). Users have also shown concern for their privacy and the possible abuse of it (Lau et al., 2018). If users are concerned with CAI and show levels of distrust towards it they will, again, be less likely to integrate it into their lives (Smith, 2018).

5. **Project outline for the work that you propose to complete**

Please complete the project outline in the box below. You should use the emboldened text as a framework. Your project outline should be between half and one page in length.

The idea for this research arose from:

The idea started because of the current situation of CAI being widespread in everyday life (Porcheron et al., 2018). People have conversations every day with

humans and as CAI usage grows in popularity, conversations with chatbots and digital assistants may occur much more frequently. Unfortunately, some people are apprehensive when it comes to using them. Whether this is down to a misunderstanding or trust issues (Smith, 2018), it is important to realise what is affecting the CUX.

The aims of the project are as follows:
Explore the field surrounding conversational AI
Research concepts such as engagement, trust and social presence
Understand what users want from conversational AI
Evaluate metrics such as duration spent with AI, how many questions are asked, and satisfaction

The main research questions that this work will address include:
What are the factors influencing interaction between people and CAI?
Are people comfortable using CAI?
Does current CAI need improving?

The software development/design work/other deliverable of the project will be:
Mixed Methods Research - users will interact with digital assistants and will be interviewed to find out what influenced their interactions with it.

The project deliverable will be evaluated as follows:
Users' experiences with different types of CAI will be compared to see how the interactions with various CAI are influenced.
Does one CAI perform better than another? If so why?

The project will involve the following research/field work/experimentation/evaluation:
Research concepts surrounding CAI
Experimentation of users interacting with forms of CAI
Evaluation of users' experience with the researched concepts
Evaluation of user metrics

This work will require the use of specialist software:
Digital assistant (such as Google Assistant, Siri, Alexa)

This work will require the use of specialist hardware:
Google Home
Apple device (iPhone/MacBook/iPad)
Amazon Echo

The project is being undertaken in collaboration with:
N/A

6. **References**
Please supply details of all the material that you have referenced in sections 4 and 5 above. You should include at least three references, and these should be to high quality sources such as refereed journal and conference papers, standards or white papers. Please ensure that you use a standardised referencing style for the presentation of your references, e.g. APA, as outlined in the yellow booklet available from the School of Computing office and http://www.soc.napier.ac.uk/~cs104/mscdiss/moodlemirror/d2/2005_hall_referencing.pdf .

Chatbots 101. (2017). [Ebook]. Retrieved 3 June 2020, from https://www.oracle.com/a/ocom/docs/chatbot-infographic-3672253.pdf.

Chatbots, a Game Changer for Banking & Healthcare, Saving $8 billion Annually by 2022. Juniperresearch.com. (2017). Retrieved 3 June 2020, from https://www.juniperresearch.com/press/press-releases/chatbots-a-game-changer-for-banking-healthcare.

Cowan, B., Pantidi, N., Coyle, D., Morrissey, K., Clarke, P., & Al-Shehri, S. et al. (2017). "What Can I Help You With?". Proceedings of The 19Th International Conference On Human-Computer Interaction With Mobile Devices And Services. https://doi.org/10.1145/3098279.3098539

Kim, K., Boelling, L., Haesler, S., Bailenson, J., Bruder, G., & Welch, G. (2018). Does a Digital Assistant Need a Body? The Influence of Visual Embodiment and Social Behavior on the Perception of Intelligent Virtual Agents in AR. 2018 IEEE International Symposium On Mixed And Augmented Reality (ISMAR). https://doi.org/10.1109/ismar.2018.00039

Lau, J., Zimmerman, B., & Schaub, F. (2018). Alexa, Are You Listening?. Proceedings Of The ACM On Human-Computer Interaction, 2(CSCW), 1-31. https://doi.org/10.1145/3274371

Lean, G. (2019). UX Best Practices to Enhance Conversational Design. Botsociety Blog. Retrieved 6 June 2020, from https://botsociety.io/blog/2019/01/ux-best-practices/.

Malkin, N., Deatrick, J., Tong, A., Wijesekera, P., Egelman, S., & Wagner, D. (2019). Privacy Attitudes of Smart Speaker Users. Sciendo. Retrieved 10 June 2020, from https://petsymposium.org/2019/files/papers/issue4/popets-2019-0068.pdf.

Moore, R., & Arar, R. (2019). Conversational UX Design: A Practitioner's Guide to the Natural Conversation Framework (p. 25). Association for Computing Machinery.
Morphy, E. (2018). What Is Conversational User Experience (UX). CMSWire.com. Retrieved 4 June 2020, from https://www.cmswire.com/digital-experience/what-is-conversational-user-experience-ux/.

Nguyen, A. (2018). Learning the basics of Conversational UI with a UX Designer for Amazon's Alexa. freeCodeCamp.org. Retrieved 10 June 2020, from https://www.freecodecamp.org/news/learning-the-basics-of-conversational-ui-with-a-ux-designer-for-amazons-alexa-c76c1908454b/.

Porcheron, M., Fischer, J., Reeves, S., & Sharples, S. (2018). Voice Interfaces in Everyday Life. Proceedings Of The 2018 CHI Conference On Human Factors In Computing Systems (CHI '18). https://doi.org/10.1145/3173574.3174214

Smith, A. (2018). Conversational UX Design : An Introduction - Usability Geek. Usability Geek. Retrieved 6 June 2020, from https://usabilitygeek.com/conversational-ux-design-introduction/
.
Statista Research Department (2019). Number of voice assistants in use worldwide 2019-2023 | Statista. Statista. (2019). Retrieved 2 June 2020, from https://www.statista.com/statistics/973815/worldwide-digital-voice-assistant-in-use/.

Sutcliffe, A. (2016). Designing for User Experience and Engagement. Why Engagement Matters, 105-126. https://doi.org/10.1007/978-3-319-27446-1_5

7. **Ethics**

If your research involves other people, privacy or controversial research there may be ethical issues to consider (please see the information on the module website). If the answer below is YES then you need to complete a research Ethics and Governance Approval form, available on the website: http://www.ethics.napier.ac.uk .

Does this project have any ethical or governance issues related to working with, studying or observing other people? (YES/NO)	YES

8. **Confidentiality**

If your research is being done in conjunction with an outside firm or organisation, there may be issues of confidentiality or intellectual property.

Does this project have any issues of confidentiality or intellectual property? (YES/NO)	NO

10. **Submitting your proposal**
    1. Please save this file using your surname, e.g. macdonald_proposal.docx, and e-mail it to your supervisor, who will discuss it with you and suggest possible improvements.
    2. When your supervisor is content with your proposal, submit it to the Research Proposal Upload link on Moodle, and email your internal examiner to notify them that you have submitted. They will leave feedback for you on Moodle.
    3. Discuss your feedback from the internal examiner with your supervisor and if necessary make final changes to your proposal.
    4. When you produce your dissertation, add your finalised proposal as an appendix.