MASTER

Identifying the Influence of Emotional Voice Style in Proactive Automobile Voice Interfaces

Nallapaneni, A.

Award date:
2021

Masters in Automotive Technology
Future Everyday Group - Department of Industrial Design

Identifying the Influence of Emotional Voice Style in Proactive Automobile Voice Interfaces

Master Thesis

Anirudh Nallapaneni
1297864
a.nallapaneni@student.tue.nl

Supervisors

Dr. Bastian Pfleging, Eindhoven University of Technology (Internal Supervisor)
Dr. Markus Funk, Cerence GmbH, Ulm, Germany (External Supervisor)
Vanessa Tobisch, Cerence GmbH, Ulm, Germany (External Supervisor)

Committee Members

dr. ir. Raymond H Cuijpers (*Committee Chair*)
dr. ir. Ion Barosan (*Committee Member*)

Eindhoven, February 2021

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Declaration concerning the TU/e Code of Scientific Conduct for the Master's thesis

I have read the TU/e Code of Scientific Conduct[i].

I hereby declare that my Master's thesis has been carried out in accordance with the rules of the TU/e Code of Scientific Conduct

Date 11/02/21

Name Anirudh Nallapaneni

ID-number 1297864

Signature Anirud

Insert this document in your Master Thesis report (2nd page) and submit it on Sharepoint

[i] See: http://www.tue.nl/en/university/about-the-university/integrity/scientific-integrity/
The Netherlands Code of Conduct for Academic Practice of the VSNU can be found here also.
More information about scientific integrity is published on the websites of TU/e and VSNU

Foreword

This is the graduation thesis report for my master's in Automotive Technology, specialising in Automotive Human Factors with the department of Industrial Design. The project was conducted under the supervision of Dr. Bastian Pfleging from Eindhoven University of Technology, the Netherlands, and Vanessa Tobisch and Dr. Markus Funk from Cerence GmbH, Ulm, Germany. Since this project was performed under the guidance of my supervisors, I have chosen to use a scientific plural throughout my report.

Acknowledgement

This project was conducted as a collaboration with the Eindhoven University of Technology and Cerence GmBH. I would like to thank my TU/e supervisor **Dr. Bastian Pfleging** for providing me with constant guidance throughout this project. I would also like to thank my external supervisors **Vanessa Tobisch** and **Dr. Markus Funk** for this giving me this opportunity and for their help throughout the project.

I am also extremely grateful towards my parents, my brother and sister-in-law for their unconditional love and support throughout my thesis. My hearty thanks goes to all my friends and colleagues, who have helped and encouraged me. Finally, I would like to express my special appreciation to **Dr. Debargha Dey** and **Melanie Berger**, whose expert feedback and motivation have assisted me greatly.

Abstract

Voice interfaces in automobiles have been gaining momentum over the past few years. Current voice systems can only be personalised by content. Emotional personalisation of voice assistants is still a relatively unexplored area. In this project, it is proposed that an emotionally styled proactive voice system would improve user experience over neutral style of voice expression. Initially, use cases were identified for an emotionally styled voice assistant. These identified use cases were prioritized and a scenario was designed based on the four prioritized use cases. A remote study was conducted with 21 participants and it was found that the emotionally styled system was rated higher in terms of attractiveness. It was also found that contextual proactivity is desired by users. This study also explored and identified other situations where voice assistants can be proactive.

Abbreviations

AI - Artificial Intelligence

CVA - Neutrally styled or Control Voice Assistant

EVA - Emotionally styled Voice Assistant

GA - Google Assistant

MBUX - Mercedes Benz User Experience

NLU - Natural Language Understanding

POV - Point of View

RF - Rank Frequency

TTS - Text To Speech

UX - User Experience

UEQ+ - User Experience Plus Questionnaire

VA - Voice Assistant

VUI - Voice User Interface

WoZ - Wizard of Oz

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Voice user interfaces (VUI) have gained popularity in devices such as smartphones, smart speakers, home automation systems and cars over the last few years. Interfaces in the form of a Voice Assistant (VA) could help users accomplish their tasks without the need for diverting their visual attention. In general, voice assistants have become increasingly capable, with abilities such as placing phone calls, voice search, navigation, making appointments, playing music, responding to texts and much more. Interaction with voice assistants became popular with smartphones (Android and iOS) and have now moved into a myriad of other devices such as speakers, headphones, televisions, watches, cars and even to home devices like security cameras. Examples of a few voice interfaces are the Google Assistant[1] found in Android smartphones, tablets, cars etc, Apple Siri[2] in iPhones, iPads and MacBooks, Amazon Alexa[3] in Speakers, Headphones, Televisions, Watches etc. and Windows Cortana[4] in laptops and tablets.

With connected technologies becoming omnipresent, car manufacturers have progressively been incorporating more methods to interact with in-vehicle technology. In the past, users interacted with in-car technologies such as radios through analog switches. In the 1980s, digital instrument panels replaced analog dials. It wasn't until the launch of the iPhone in 2007 [5], that multi-touch interactions gained popularity. Touch graphical interfaces have since become commonplace, with almost all automotive manufacturers of today incorporating multi-touch screens in car dashboards or instrument clusters. A few infotainment systems allow users to interact using touch as well as gestures as shown in Fig 1.1.

Interaction with a screen in the instrument cluster or on the dashboard requires drivers to divide their visual attention away from the road. Most modern infotainment systems have multiple interaction styles such as touch input, buttons, dials, voice and gestures, which can be used individually or together.

Digital touchscreen interfaces have come to replace analog controls and displays. However, driving is a task that demands visual attention of the user. Drivers could be at higher risk when their visual attention is moved away from the road, to visual interfaces inside the car. Voice systems could work as a good alternative for drivers to interact with the technology while keeping their visual attention on the road. In addition, voice interaction could also help users accomplish tasks faster when compared to visual interfaces in a few specialised situations. For example, if a user had to play a specific song from their touchscreen, they would have to go into their music library. The next step would be to click on the search bar, and then typing out the song to be played. Finally,

[1] https://assistant.google.com/

[2] https://www.apple.com/siri

[3] https://www.amazon.com/b?ie=UTF8&node=17934671011

[4] https://support.microsoft.com/en-ca/help/17214/cortana-what-is

[5] https://tinyurl.com/y5vuvqw9

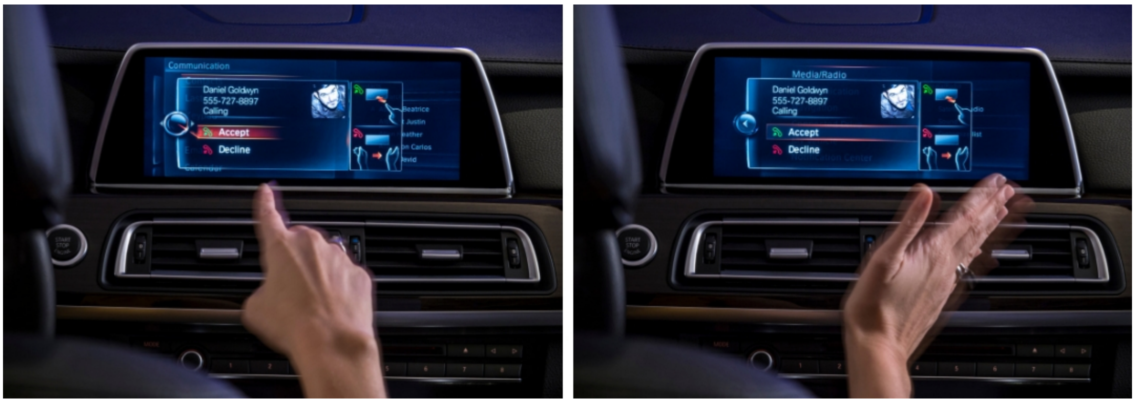


Figure 1.1: Gesture control in the BMW 5 series[6]

the user has to click on the song for it to start playing. A process such as this, takes about four steps to complete. In contrast, for voice interactions, the natural language understanding system in system matches the voice prompt request to the action of playing the song. From the user's perspective, it could be as simple as prompting "Play Blinding Lights by the Weeknd" to the system. The same task can therefore be completed in one or two steps via voice in comparison to the four steps required for a touchscreen interface. Therefore, voice interfaces could potentially be a safer and faster replacement for specific tasks, when compared to visual interfaces during driving.

In the automotive space, there are a plethora of voice interfaces specifically designed for in-car interaction. Examples such as, Android Auto an extension of the Google Assistant, Apple CarPlay which uses Siri by Apple, and Echo Auto by Amazon Alexa are specialised automotive interfaces with voice functionality. These three interfaces are essentially extensions of the systems that exist over multiple devices. Therefore, these systems could potentially have access to user data such as contacts, music library and calendar etc. which allows personalised content to be presented to users. Content personalisation could help improve the interaction with users as it would provide specifically tailored interaction methods to users.

Furthermore, voice assistants can be categorized into cloud based or local assistant depending on the need for an active internet connection to carry out functions. Cerence GmbH, the client for this project, creates embedded voice assistants in collaboration with automakers. Cerence specializes in making local interaction systems with a focus on voice interaction. They develop dedicated proprietary voice assistants that are built into the cars' user interface systems. Examples of such systems are the Mercedes Benz User Experience (MBUX), BMW Connect, FordSYNC etc. With high customer demand for cloud based voice services in their cars, automakers slowly start implementing multiple voice systems [52]. For example, in the MBUX voice assistant, a hybrid voice control system allows users to interact even without an internet connection by using both an in-car assistant and a server-based assistant to find an optimal answer to user requests [7]. Another example is the BMW intelligent personal assistant which acts like a co-driver. It can help the driver with various functions from remembering preferences and providing casual conversations to complex tasks such as triggering the vitality program that adjusts mood lighting, music and temperature when the driver says that they are tired [34]. Along with these proprietary systems, some cars have a combinations of Android Auto, Apple CarPlay and Echo Auto. This enables car companies to cater to various preferences of multiple users.

Current voice interfaces in cars now are mainly trigger-based systems i.e. they require the user to press a dedicated button or to activate the interface using a wake-up command such as "Hey

[6] https://www.bmwblog.com/2019/11/10/video-how-to-get-the-most-out-of-gesture-control/

[7] https://www.daimler.com/magazine/technology-innovation/mbux-voice-assistant-hey-mercedes.html

Mercedes", "Hey BMW", "Alexa", "Hey Siri" etc. These systems are limited in the sense that the interaction can only be started by the user and the system responds to this request. Proactivity of such systems could occur by creating or initiating an interaction rather than responding to it. In the near future, cars having shared control with drivers as seen with Level 3 and 4 automated vehicles [38] will become more commonplace. The driver has to transition control from and to the vehicle. In such situations, a proactive VA could enable users to have a natural conversation with the VA taking the role of a co-driver, rather than being a disconnected system that is not contextually aware of its surroundings. Proactivity also opens up a multitude of new functionality for the VA with contextual and relevant information being presented to the user without the need explicit user commands.

Human to human communication is an intricate process which involves more than voice communication. Humans communicate through body language, facial expression and emotional expression within their voice. To enable natural interaction between a voice assistant and a human this research tries to investigate emotional expression by the system. Therefore, the primary research goal for this project was to **investigate the effects of varying emotions using voice styles in a proactive voice interaction system**. By introducing emotional styling into the voice expression of an assistant, the goal is to improve the user experience of interactions with VAs. Emotional voice styling is the first step towards the ultimate goal of improved interaction with technology with users. The emotionally styled VA used in this project only uses two fundamental dimensions of emotions i.e. negative and positive. The goal of this project is congruent with our client's goal of improving personalised voice interactions for their customers.

As a starting point for the current project, user requirements were collected through surveys, brainstorming sessions and expert reviews. These requirements were used to create scenarios for the user study. A conversation was designed and a total of 21 participants were recruited for the user study. Both quantitative (survey) and qualitative (interview) feedback was collected for the proposed intelligent voice assistant. Finally, based on the data analysis from the user study, a few conclusions were drawn and future work along with limitations were discussed. It was found that users rated an emotionally styled voice assistant (which matched emotions according to user expectations) better, than a neutrally styled voice assistant. Other situations where users desire VAs to be emotional and proactive were identified. User perception towards proactivity was investigated and it was found that contextually proactivity could be desired highly by users.

The related work chapter of the report is a discussion of the theoretical background through literature research. After that, the next chapter explains the user centered design process that was followed during the project. Further on, we discuss our conceptual idea and detail our prototype in the concept chapter. We then explain our experimental setup for the user study and detail the procedure in the chapter titled experiment. Further on, we present our results and discuss their implications. Finally, we draw conclusions in the last chapter based upon the results and observations from the user study. We also talk about the limitations of this study and future work.

Chapter 2

Related Work

In this chapter the background and related research with reference to this project is discussed. Related research was used to narrow down the scope of this study and to use it as a structure to support the framework on which this project is based.

2.1 Voice Interaction in Automobiles

With the number of smartphone users on the rise[1], a capable digital voice assistant(VA) could potentially allow users to interact with infotainment systems while eliminating the need to look at visual interaction devices while driving. Systems such as Apple CarPlay, Android Auto and Alexa Auto enable users to mirror their smartphones with limited functionality onto their car infotainment screens. Some of the functions are navigation, music, podcasts and messages (via voice).

Early speech recognition systems used acoustic based systems. The system would try to match the sound of a spoken dialog as closely as possible to a set of pre-programmed tonal database. However, it was found that acoustically based systems were limited in their functionality [10]. There was a shift towards a linguistic approach, which involved using algorithms to program systems with rules of a language. Linguistic modelling is the basis behind a modern Voice User Interfaces (VUI).

The principle behind how a modern VUI works is shown in Figure 2.1. The user's speech signal is recorded by the speech input system. Speech input is then decoded using a decoder. The decoder converts the recorded data into text by using Natural Language Understanding (NLU). NLU uses acoustic, pronunciation and language modelling which allows VUIs to structure and transform human language into a machine readable format. After understanding the meaning of the human prompt, an appropriate response is chosen by the voice app based on artificial neural networks. A text-to-speech (TTS) system then converts this text into a voice output and plays the voice command to the user. Improvements in speech recognition technology, text to speech and the use of machine learning has allowed VUIs to be capable of carrying out complex conversations. These advances in VUI technology have enabled users to use natural conversation for interaction. A gain in popularity of VUIs in smartphones, smart-speakers, smart-home systems, and other devices was seen because of such advances.

The use of VUIs has increased over the past few years with nearly one third of U.S adults being active users (atleast once a month) of in-car Voice Assistants(VA) in 2020 [52]. In this Voicebot

[1] https://tinyurl.com/y4vu5wkf

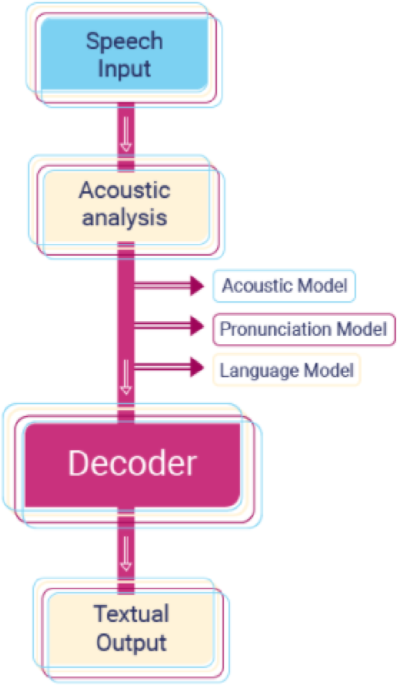


Figure 2.1: Working of a Voice User Interface[2]

report, it was also projected that the number of users of voice systems in automobiles will increase over the next few years. Multiple studies have shown that voice is a suitable alternative to visual interfaces [48],[23] and [32]. In a study [48], users rated voice interfaces as easier to use and more satisfying, when compared to visual interfaces in cars. Visual interfaces were also rated as less efficient as it strongly distracted the user from performing the primary task, i.e driving [48]. David et al. found that voice interaction requires low to medium cognitive load during driving [23]. In this study, David et al found that the cognitive demand required for a driver to interact with a digital driving VA was similar to that of having a hands-free mobile conversation. Voice interaction was found to be a superior alternative in terms of driver performance, distraction and accident hazards when compared to a graphical or visual display [32]. Therefore it is safe to assume that voice interfaces could be a suitable alternative to visual interfaces while driving.

However using an auditory interface does come with drawbacks of its own. When the voice system does not work according to the user's expectations, it was found that driving performance was lowered [22]. An accurate speech recognition system along with an easy to use voice interface are needed to improve driving performance and reduce accidents according to a study [32].It was found that interacting with a state-of-the-art digital driving assistant had higher physiological arousal, subjective workload and lower driving performance when compared to the baseline of no secondary task [23]. This suggests that voice interaction might reduce the driving performance when compared to no interaction at all. However, it was seen that only 1 in 5 drivers automatically set "Do Not Disturb While Driving" mode on their smartphones [36]. This could suggest users are reluctant to stop carrying out secondary tasks while driving. Therefore, voice interaction can potentially be a safer alternative to visual interaction, for carrying out secondary tasks while driving.

[2] https://idapgroup.com/blog/basics-of-voice-assistants-explained/

2.2 Proactivity

Devices such as smartphones,tablets and laptops alert users through the use of notifications. Such notifications could be classified as being partly proactive. A proactive system should know the right moment as well as the right information to present to users. A smartphone device only notifies users when a prompt is received, either when another person sends a message, or via an application which receives an alert. Such devices automatically alert users with notifications whenever they are received. Users have the explicit choice to choose what kind, what time and how they receive notifications. Such explicit choice could give the system instructions on how, when and what to be proactive about.

Notifications can be multimodal i.e consisting of earcons (a brief distinctive sound meant for a particular event) as well as a visual message on the display of the device. These notifications could sometimes occur at the inappropriate time and hence annoy the user. In the field of social robotics Weber et al. propose designs to make such notifications less disruptive to the end user [57]. Other proactive interactions through social robots include activity by detecting the presence of the human based on contextual information (e.g Invitation to play a game, telling a joke etc)[3]. Though this robot cannot physically help the human with their task, it can entertain them with some quirky features. The Anki Robot which can be called as a physical embodiment of a virtual assistant that is capable of being proactive. It was proposed that user perception of the robot in terms of personality, emotional and social intelligence also plays a role in it's proactivity [9]. A proactive system should try to minimize intrusiveness when interacting with the user and must interact only when it seems “right”.

In the automotive field, empirical research was conducted to find this “right time” to be proactive for a voice interaction system. A research study focuses on driver preferences and self-assessment of availability to measure an opportune moment for a proactive interaction [46]. In this real world driving test, it was found that people would not want to be interrupted by non-critical interaction when they are driving off-course or lost [46]. Semmens et al. suggested that drivers have a high likelihood of wanting to be interrupted while driving in a straight line or waiting at a signal [46].The results of a survey showed that participants desired VA that could be proactive [42]. Schdmidt et al. also hypothesised that the acceptance threshold of such proactive systems may vary with type of users and type of notifications. Users also desired personal assistants in a vehicle to be adaptive towards situations, make proactive suggestions and expressed the willingness to share personal data, if it serves the needs of the user in [42].

In the automotive context, providing wrong information could not only increase annoyance for the driver, but could potentially compromise safety. To identify such situations, Schmidt et al. explored a few use cases for in-car proactive VA [44] and observed that proactivity of the specific driving use-cases such as proactive gas refueling, car parking and break management had a high acceptance by users. Researchers found that proactive VAs are at least equally likeable as non-proactive assistants while driving [43]. Schmidt et al. further goes on to say that even though a few users would like to disable proactivity for specific functionality. These findings are in agreement with those found in other research [42]. The majority of users rated proactivity positively in a high traffic situation within a controlled environment i.e a driving simulator. This suggests that proactivity could be a context dependent feature, which needs to be personalised to user preferences. In contrast, it was found that a majority of users opposed unsolicited speech in [8]. However, Braun et al. found that if unsolicited speech was found appropriate by users, they preferred to be interrupted by personalised characters.

[3] https://anki.com/en-us/vector.html

2.3 Personalisation of Voice Interfaces

The idea behind personalising digital systems is to try and provide a tailored experience for particular individual user [18]. Such personalisation is possible by making use of user information or content to provide recommendations based on their preferences. Currently, the available VAs make use of such user data, to provide a personalised interaction. Another way of personalisation could be to develop a VA personality which matches the personality of the user. However, current VAs in the industry use a standard personality which are not customized to individual users. The similarity-attraction hypothesis proposes that users like to interact with similar personalities as themselves when communicating with both humans and computers [31].

A two dimensional model was used to identify personality dimensions during interactions with digital agents in a study [6]. The first dimension ranged from equivalent to subordinate and the second dimension from formal to casual. Braun et al. used the 2-D model to develop various VA personalities [8]. They used the similarity-attraction hypothesis to propose that personalisation matching of VAs, provided better trust and likability of the system by users. However, this increased trust and likability was only for systems that had correct personality matching. They found that the personality matching proved to be challenging, as the algorithm they used matched only 29% of the participants correctly. Therefore, it was proposed that implicit decisions made by the system may not always perform as expected and explicit user input could be used to better match the personality of users to that of the VA. Furthermore, they recommend that the VA for non-driving-related situations should be personalised, but for driving related functions it should be unemotional.

A conversational agent should try to provide good user experience, so that user satisfaction improves while interacting with such systems. Neuroticism, a Big Five Model trait, could be seen as a negative, stressful or moody behaviour and could potentially reduce the user experience of the conversational agent. It was seen that the VA with neuroticism was rated poorly by users [8].

In order to create acceptance for such affective interactions, already existing features should be given emotional components according to [7]. In the field of social robotics, it was found that personality matching was preferred by users and also had positive effect on the motivation in socially assistive robots in [2] and [50].

2.3.1 Emotional Expression by Voice Assistants

Watson et al. conceptualized positive and negative affect as the two fundamental dimensions of an emotional experience [56]. Emotions can be clustered positive and negative emotions into two categories [27] as shown in Fig 2.2. As a first step towards natural interactions with VA, the emotional VA used in our project only uses these two fundamental dimensions of emotions.

It was found that matching a car's voice interface emotion to driver emotion, improved the driving performance and safety [29]. Nass et al. explored driving related tasks to improve driving safety, by pairing voice and driver emotions in a car [29]. In a long term study, it was found that relational agents capable of displaying social-emotional skills were rated highly for trust and likability when compared to a task oriented agent [4].A relational agent, which is defined computational system was built to establish long term, social-emotional relationships with users [4]. Users also expressed a desire to continue working with such a relational agent after the end of the study. It was found that 80% of users would generally accept emotion based mood improving voice interaction, which suggests that a VA capable of showing emotions could be perceived positively [3]. The situations where people would accept mood improving services are for relaxation, mediation, motivation and fun and entertainment [3].

[3] doi = 10.1109/T-AFFC.2013.25

According to research, drivers exhibiting negative emotions, such as anger, can increase the likelihood of accidents [51]. Zepf et al. found that triggers associated with traffic and driving task frequently evoked negative emotions [58]. On the other hand, the vehicle and its equipment invoked positive emotions. Finally, they propose interventions for better regulation of drivers emotional state, in order to increase safety.

The appraisal theory claims that emotions are elicited and differentiated based on the subjective evaluation of the personal significance of the situation [41]. They further go on to say that individual evaluation causes an emotional or affective reaction to a particular situation. Therefore, a same event or situation could elicit variable reactions from different people according to this theory. The appraisal theory is in contrast to the personality psychology that was used as the basis for [8], [54],[42] and [6].

Paul Ekman identified six basic emotions that are used during communication by humans and they are namely; Happiness, Sadness, Fear, Disgust, Anger and Surprise [14]. But the classification was made for human to human communication which involves facial cues to display emotions. A study further supports this [47], which found that a VA with facial emotional expressions or text body movements can evoke stronger user engagement when compared to voice waveform emotional expression. However, an interaction with a VA should not contain visual modality, especially within the automotive context as mentioned in the previous section. Therefore, for the scope of this study, voice style modulation was selected as an initial step towards understanding user perceptions towards emotional VAs.

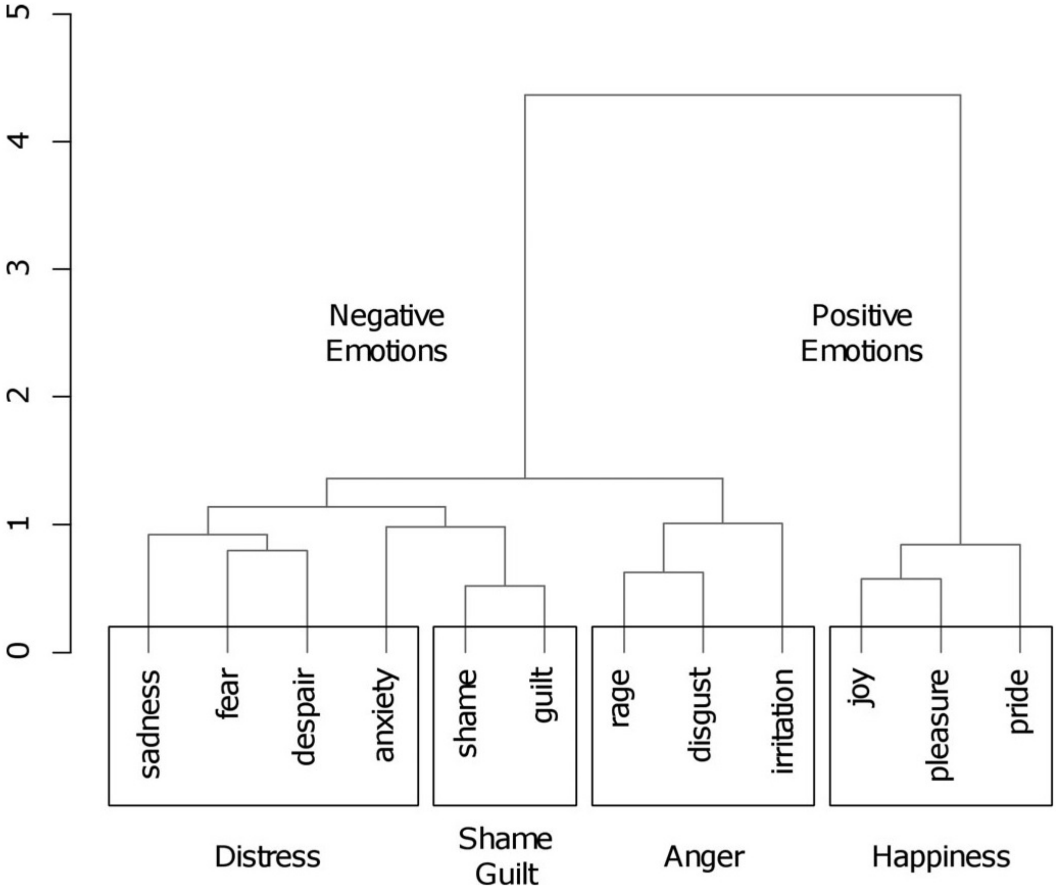


Figure 2.2: Clustering of emotions[4]

A manipulation check was conducted in the current study, to identify user's subjective evaluation towards the emotional voice system. The reason behind this was to check whether the the desired emotional output of the VA matched the subjective evaluation of that voice prompt by the user i.e. to check if users perceived the negatively styled emotional voice as a negative emotion and positively styled as a positive emotion.

2.4 Research Direction

Previous research investigates using personalised voice characters for a proactive use case in automobiles by matching driver and assistant personalities [8]. Other research explores the timing of when a system can be proactive [46]. Driving related tasks were explored to improve driving safety by pairing voice and driver emotions in a car[29]. Various use cases for proactivity were explored and assessed in [44]. Our project is unique as the area of investigating proactive voice systems by using emotional voice styling for non-driving related activities is relatively unexplored. We are building on the works of Braun et al [6] to establish the emotional dimensions to explore proactivity of voice interfaces.

For the current project, we explore whether introducing emotions in a proactive voice assistant improves user experience. By using such a proactive VA capable of displaying emotions, we hope to build a connection between the driver and the assistant. Such an attachment could make voice interfaces more likable by customers and potentially improve brand loyalty for our client, Cerence GmbH.

Chapter 3

Design Process

In this chapter we discuss the background research conducted for our project. We implemented a user centered design approach for our project [1]. We conducted multiple user studies such as market research, surveys and focus groups during the course of this project. We used an iterative process by evaluating the data collected from these studies and conducting further user research. This user data, along with the findings from related work, helped define a coherent scope for this project.

To reduce complexity in the report, all the user studies have been grouped according to the method, i.e. focus groups, surveys, expert review etc. However, the user centered design process involved iterative methods of conducting user studies. Therefore, the chronological order of the tasks that were carried out during this study is explained on the next page in Fig 3.1.

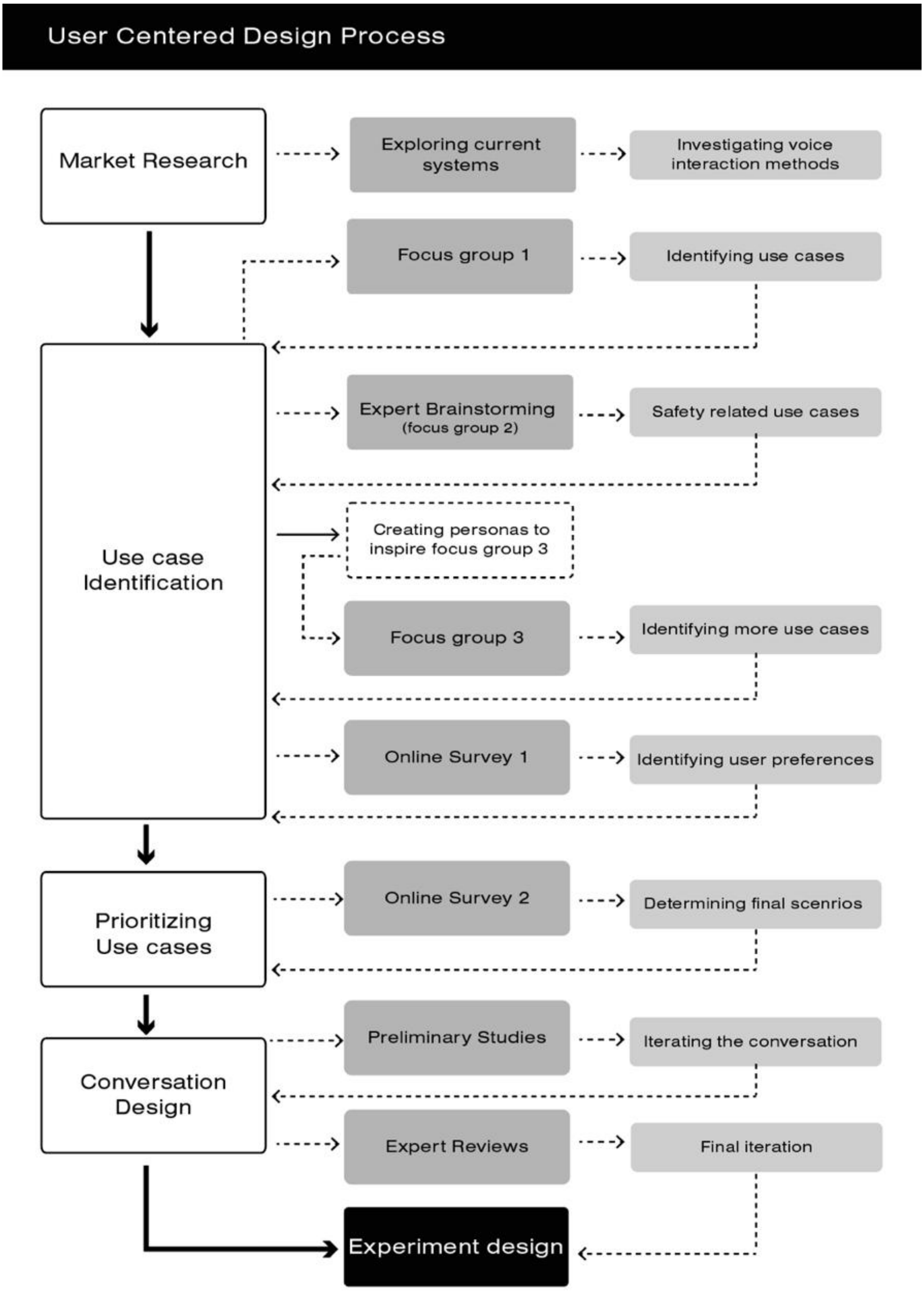

User Centered Design Process
Market Research
Exploring current systems
Investigating voice interaction methods
Focus group 1
Identifying use cases
Use case Identification
Expert Brainstorming (focus group 2)
Safety related use cases
Creating personas to inspire focus group 3
Focus group 3
Identifying more use cases
Online Survey 1
Identifying user preferences
Prioritizing Use cases
Online Survey 2
Determining final scenrios
Conversation Design
Preliminary Studies
Iterating the conversation
Expert Reviews
Final iteration
Experiment design

Figure 3.1: User centered design process framework

3.1 Market Research

As the first step for narrowing the scope of this project, we conducted an extensive market research with existing voice interfaces. We identified four popularly used voice systems, namely, Google Assistant, Apple Siri, Amazon Alexa and Microsoft Cortana. A comprehensive investigation was carried out by using devices enabled with each of these systems. The purpose of this was to identify existing use cases about daily updates (proactivity) and to identify use cases with emotional styling. Apart from physical testing these VAs on smartphones and other devices, we studied the manuals and websites of these systems to identify how and where interactions take place within each VA. We examined an Alexa enabled smart speaker shown in 3.2 along with the Alexa application. However, we explored Cortana, Siri and Google assistant with the applications on a smartphone and a tablet. Ideally, we would have liked to explore these interfaces in an automobile, but due to the current pandemic this was not possible.

3.1.1 Microsoft Cortana

Cortana is a proprietary voice assistant developed by Microsoft. It utilizes Microsoft's platforms such as Azure and Office 365 to provide connected car platform for automotive manufacturers such as Nissan and BMW[1]. Cortana can be integrated with a user's office 365 account, to enable calendar access to the VA. Functionalities such as daily briefing email and play my email allows users to organize their schedules and emails in a voice forward way. The daily briefing email provides an overview of meetings and tasks with a plethora of customizable options. During the setup of daily briefing, the system explains how to use it to prep for future meetings and manage pending tasks. Users are given the option to interrupt or dismiss the briefing at any time. Daily briefing only starts on explicit user command, therefore it cannot be said to be truly proactive. However, it makes use of calendar data to identify scheduling conflicts, task briefing and tries to provide relevant documents for meetings proactively. To summarize, even though the voice system does not proactively initiate conversation, the system is capable of using user data to provide proactive suggestions to mimic the behaviour of a human assistant in a voice forward manner.

3.1.2 Apple Siri

Siri is a voice interface that exists over different devices such as the iPhone, iPad, Laptops and in automotive interface under the name of CarPlay[2]. Siri is primarily a reactive VA which can answer a variety of user queries, from playing music to telling a joke. CarPlay allows users to read, write listen and reply to messages using voice commands with Siri. Other functionalities include navigation, music/podcast control and making phone calls. Siri can also provide personalised shortcuts for weather, news or other topics.

3.1.3 Google Assistant

Google Assistant(GA) exists in multiple devices such as phones, speakers, laptops and cars in the form of Android Auto[3]. Similar to Siri, Google assistant has multiple functionalities. With respect to proactivity, GA allows users to subscribe to a daily updates from its "actions" menu. Actions allow users to extend GA functions through a development kit. Multiple scenarios can be triggered from actions and other companies that collaborate with GA can develop an action

[1] https://tinyurl.com/y5klk45p
[2] https://www.apple.com/siri/
[3] https://www.android.com/auto/

specific to their context. Some developers have configured the Tesla app to work with GA which enables users to carry out tasks such as finding out charge status, locating their car, turning on the climate control system etc from the comfort of their home, using any GA enabled device.

3.1.4 Amazon Alexa

Amazon Alexa is a voice assistant developed for all kinds of devices from smartphones, speakers, televisions and cars (Alexa Auto)[4]. Alexa allows developers to add their "Alexa skill" in addition to existing functionality. Alexa also gives a daily flash briefing which includes content such as news, traffic, comedy and interviews. Alexa auto lets users search for parking, call emergency assistance, make dinner reservations, control devices in smart home, order at restaurants when the user is close. A recent addition to Alexa skills that was developed during the pandemic, allows users to make contact-less voice payments at fuel stations.

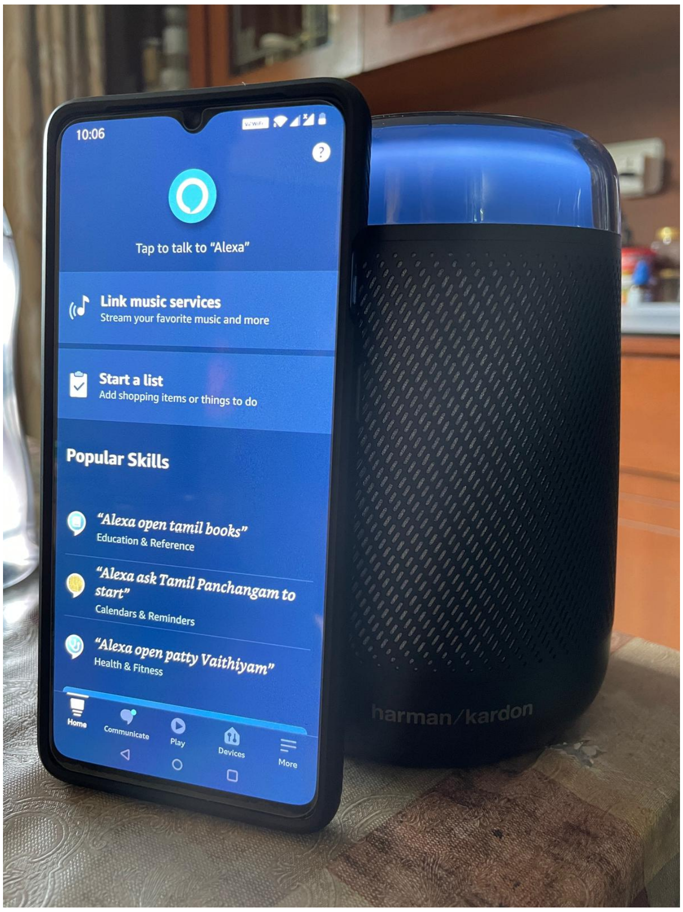


Figure 3.2: Alexa enabled smart speaker

3.2 Lessons Learnt from Market Research

One particularly interesting Alexa skill we found was the "Quiz of the day". Quiz of day lets users play a general knowledge quiz using voice interaction with Alexa. In this skill, we noticed that

[4] https://www.amazon.com/alexa-auto/b?ie=UTF8&node=17599297011

Alexa uses an emotionally styled voice expression. Alexa praises users when they answer correctly by using a happy voice style. On the contrary, it uses an empathetic voice with users, when their answer is wrong. We also noticed that, Alexa uses sarcasm a few times to make a joke when user answers incorrectly to the questions in the quiz. This varied voice styling of Alexa inspired us to explore this area of using emotions voice expression in a VA and this discovery was used as the starting point for our project.

Another inspiration we took came from the calendar functionality by Windows Cortana. We used this idea to prime users that the VA could have access to their calendar data and can help them with task related to scheduling. This led us to identify more use-cases or scenarios where emotional behaviour could be acceptable for users. Therefore, we decided to explore this area of affective computing with a focus on emotional styling of voice expression.

3.3 Preliminary User Research

The initial step of a user centered design process is user analysis and research [1]. To understand the context of use of VAs, we collected user data in the form of multiple surveys, focus groups and expert reviews, which will be elaborated in further sections. The data gathered from these preliminary studies helped us find and prioritize the most relevant use cases and gave us initial feedback from potential users. It also helped us define the scope of this project as well as to propound a few specific research goals and objectives.

The core part of a design process starts with identifying contexts where users could potentially interact with an intelligent voice assistant. To identify potential use cases, we explored related research in this field. In a study, four driving related use cases were found where VAs can be proactive [44]. The four use cases are:

1. Proactive Gas Refueling
2. Proactive Break Management
3. Proactive Car Parking
4. Suggestions along the route

In this study, proactive parking got the highest acceptance rate and additional value score [44]. Schdmit et al. propose that driving related use cases have high potential for customer satisfaction.

A total of 20 affective use cases were developed and evaluated in research conducted by Braun et al [7]. These use cases were split into categories of navigation, proactivity, family, control, sharing and transparency. Navigation, family and proactive use cases were rated well in hedonic and pragmatic quality by the participants of this study [7].

3.4 Focus Groups

We conducted three focus groups to identify use cases and collect user inputs on how they use their voice interfaces. The identified use cases were categorized based on common themes. Due to the pandemic, all the focus groups were conducted online via Microsoft Teams or Zoom.

3.4.1 Focus Group to Identify Use Cases

A focus group with three people was conducted in this session. The participants of this session were all masters students in the field of automotive human factors. The participants for this focus group were recruited from the first survey.

Goal

The goal of this study was to explore the interaction between voice assistants and users within an automobile. Participants were asked to think of situations where a VA can be proactive. The duration of this session was about 60 minutes.

Procedure

The primary researcher introduced themselves, welcomed the participants to the focus group and informed the participants about the goal of the focus group. Participants were then asked to give their consent for the use of their inputs within the focus group, see Appendix A.2. They were also informed that their audio would be recorded for transcription. After receiving consent, participants were sent a link for a shared online word file.

The brainstorming session consisted of two parts. The first exercise was individual ideation, where participants had to think of situations where an intelligent voice interaction takes place. The second exercise included interactive clustering of ideas using an online tool[5].

Exercise 1:*Participants were given the following instructions for this exercise*:

"Voice assistants at this moment are reactive. They require user input; either through a wake-up command or via a physical button. This limits the usability and user experience of such interfaces. Future voice assistants will have to shift from being reactive to proactive. Proactivity can be described as anticipatory, change-oriented and self-initiated behavior. In cars, voice interfaces have shown to reduce cognitive load when compared to visual interfaces. The use of voice interfaces can therefore help reduce accidents due to driver distraction when compared to interactions with visual interfaces. What are some daily situations or scenarios that you can think of where the voice assistant in a car can be proactive? An example could be the assistant giving you sports update about your favorite team. Also think about how an affective system can implement this proactivity. For example, if your team loses a game, the assistant could convey this message in a sad tone. Try to be creative, there are no right or wrong answers. What are some very specific things that you would like your car to be proactive about? Write down your responses in the pages below. You can discuss your ideas and have about 15-20 minutes for this exercise. I will notify you when time is up, let me know if you are done earlier."

After finishing the first exercise, participants of the focus group were provided with a link to a shared Figma platform. The individual ideas from the shared word document were copied onto a

[5] www.figma.com/

shared work-space on Figma. Participants were given brief instructions on how to use Figma.

Exercise 2: Participants were then asked to explain and present their own ideas to the rest of the group. Each idea was deliberated and discussed by the whole group and subjected to thematic analysis. New ideas were placed into existing themes or a new theme was developed if the members felt the need for it. After grouping the individual ideas into themes, the moderator (primary researcher) of the focus group read through all the ideas and discussed with the group if the theming was appropriate. This exercise lasted for about 25 minutes.

Results

The individual ideas were group into themes for the final categorization as shown in Figure 3.3. A total of 24 use-cases were collected in this session and they were grouped into 7 categories. A few ideas from this focus group were suggestions on the implementation of proactive VAs. Since these ideas cannot be classified as use-cases, they were omitted from the final list as seen Appendix A.1.5. The most important takeaways from this that scheduling and contextual proactivity had the highest number of ideas in its cluster. This suggested that users might want to complete their scheduling tasks while they are driving.

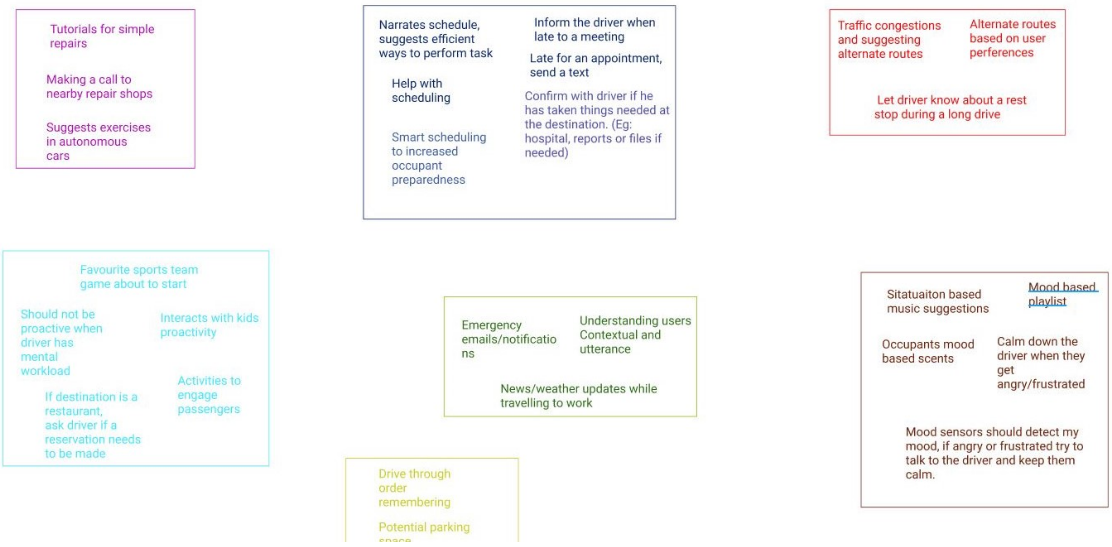

Tutorials for simple repairs
Making a call to nearby repair shops
Suggests exercises in autonomous cars
Narrates schedule, suggests efficient ways to perform task
Inform the driver when late to a meeting
Late for an appointment, send a text
Help with scheduling
Confirm with driver if he has taken things needed at the destination. (Eg: hospital, reports or files if needed)
Smart scheduling to increased occupant preparedness
Traffic congestions and suggesting alternate routes
Alternate routes based on user perferences
Let driver know about a rest stop during a long drive
Favourite sports team game about to start
Should not be proactive when driver has mental workload
Interacts with kids proactivity
Activities to engage passengers
If destination is a restaurant, ask driver if a reservation needs to be made
Emergency emails/notificatio ns
Understanding users Contextual and utterance
News/weather updates while travelling to work
Sitatuaiton based music suggestions
Mood based playlist
Occupants mood based scents
Calm down the driver when they get angry/frustrated
Mood sensors should detect my mood, if angry or frustrated try to talk to the driver and keep them calm.
Drive through order remembering
Potential parking space

Figure 3.3: Clustered ideas from first focus group for use-case identification

3.4.2 Expert Focus Group

The procedure followed for the second focus group was similar to that of the first brainstorming. This session consisted of three participants from the adaptive mobility squad at TU/e. The participants consisted of a PhD student, a master graduate and a Postdoc researcher.

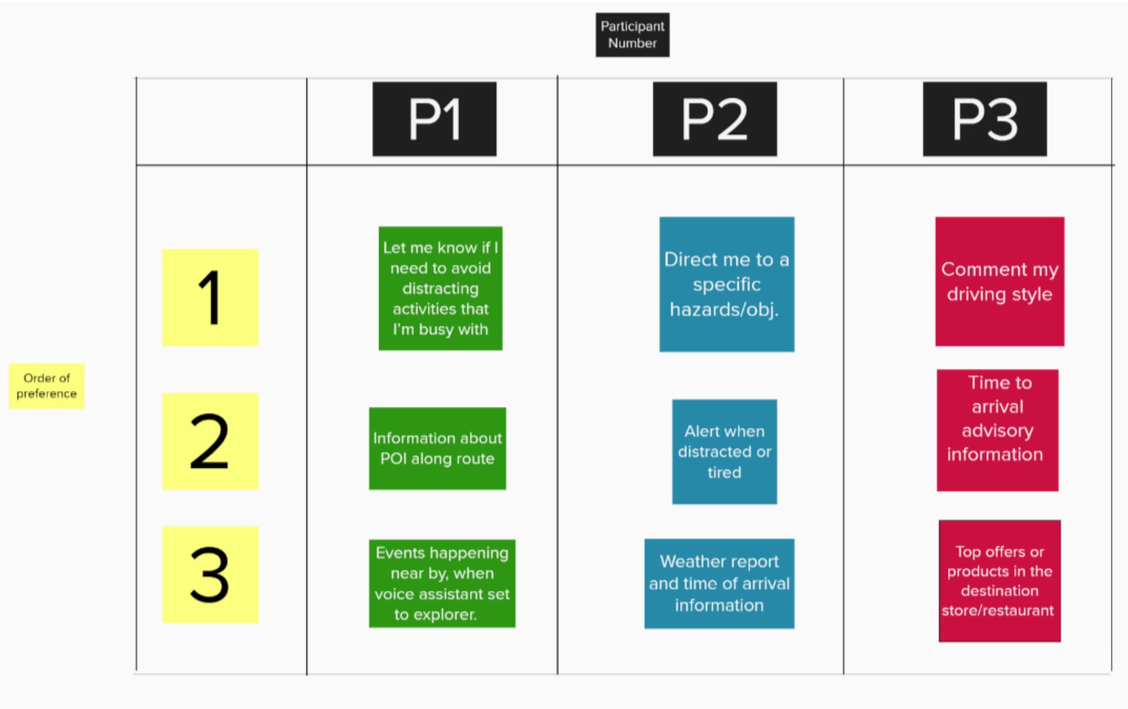
Goal

The goal of this focus group was to identify scenarios where a VA could potentially show emotions. Participants were also asked to brainstorm about use cases for a proactive system.

Procedure

For this session, we decided to use another platform known as Mural[6] for collaborative clustering, over Figma. Mural was chosen because it had better functionality and was intuitive to use, with a provision to generate and move around sticky notes over the board. Participants were asked to complete the same exercises as from the previous focus group 3.4.1. The duration of this study was about 60 minutes.

After finishing both the exercises, participants were asked to pick their top three preferences from all the use cases. The top 3 preferences from the expert focus group are shown in Figure 3.4. The idea behind asking participants to do this was to try and prioritize the use cases.



Participant Number

Order of preference

	P1	P2	P3
1	Let me know if I need to avoid distracting activities that I'm busy with	Direct me to a specific hazards/obj.	Comment my driving style
2	Information about POI along route	Alert when distracted or tired	Time to arrival advisory information
3	Events happening near by, when voice assistant set to explorer.	Weather report and time of arrival information	Top offers or products in the destination store/restaurant

Figure 3.4: Top three preferences from expert focus group

Results

The final results from brainstorming session with experts are shown in Appendix A.1. We noticed that during this brainstorming session, the experts explicitly said that safety was their major concern when picking a use case for voice interaction. It can be seen from Figure 3.4 that the top use cases by experts in the automotive human factors field include only those related to the safety of driving. In retrospect, the instructions for the focus group should have explicitly mentioned that we were also looking for non-driving related activities. Therefore, other than the use case to brief drivers about the weather at the start of their ride, the responses received from this focus group were not as useful as we expected.

[6] https://mural.com/

3.4.3 Focus Group with Potential Users

The results from focus group with experts revealed that their ideas would be limited, with safety as their primary concern. We did not want to restrict the creativity in ideating scenarios. Therefore we decided to conduct another focus group with 4 participants who were within our potential target user group as mentioned in Section 5.2.4. The participants from this focus group were recruited from the first online survey.

Goal

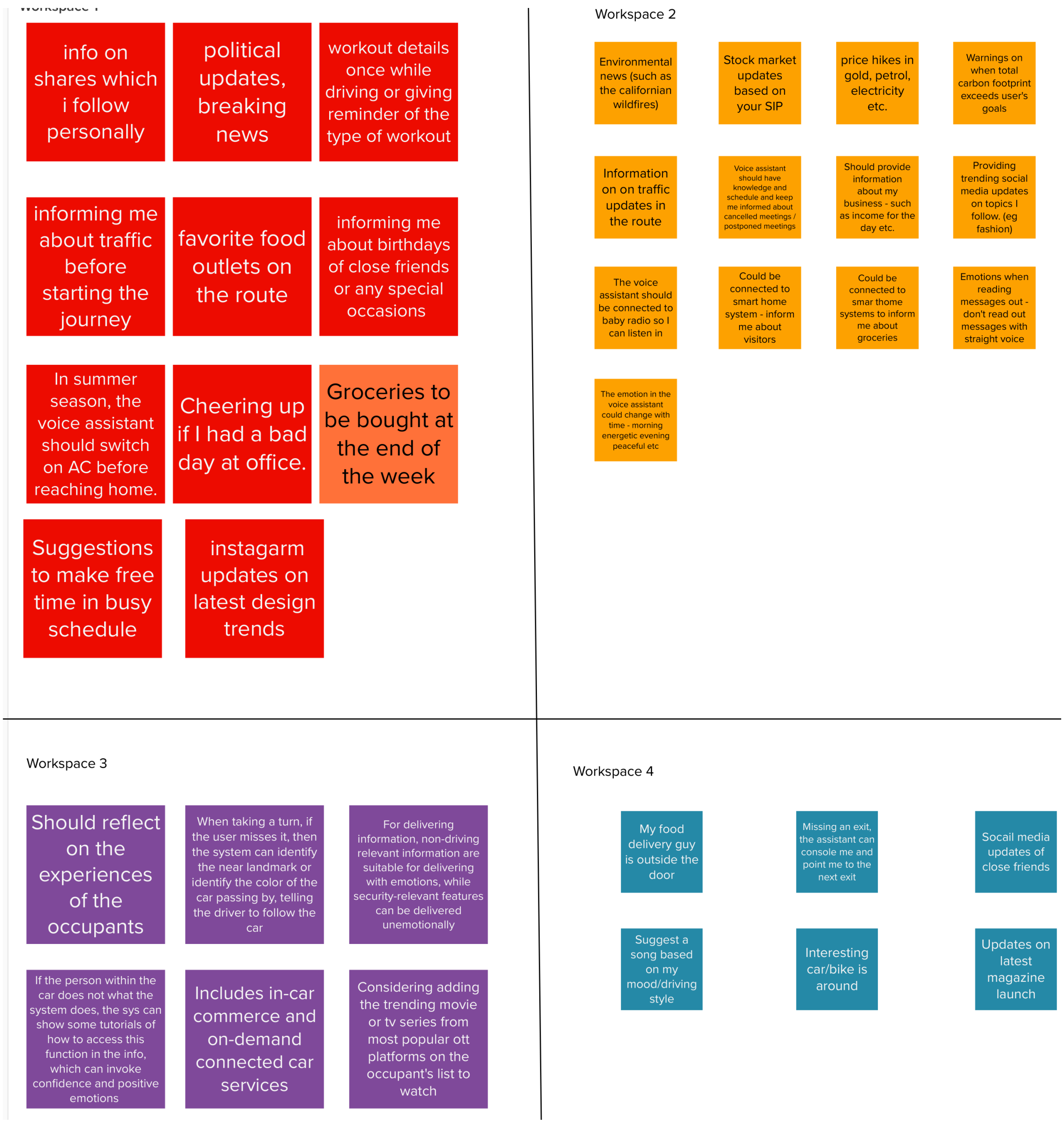
The goal of this session was to identify more non-driving related use cases for voice interactions within automobiles, from the perspective of potential customers of Cerence.

Procedure

To identify more non-driving use-cases we developed three personas as shown in Appendix A.1.6. These personas were created using the data collected from the first two brainstorming and the first survey. These personas were used to act as inspiration to participants of this session. Participants were asked to put themselves into the shoes of the personas and explore suitable use-cases from a different perspective. The first exercise in this focus group involved individual ideation. Participants were asked to think of scenarios where a VA can show emotions and they were given 25 minutes to write down all their ideas. They were also given access to a Mural work-space which included three personas to use as inspiration for their ideas. The second exercise involved clustering ideas based on themes in the Mural work-space. The total duration of this session was 1 hour and 15 minutes.

Results

The results from this session are shown on the next page in Figure 3.5. The use of personas as a method to influence user perspectives for identifying scenarios for intelligent voice interaction resulted in creative ideas from this session. Together with the use cases identified in the first two brainstorming sessions and the first online survey, we collected a total of 4 use cases for voice interactions as shown in Appendix A.1.5.



Workspace 1

info on shares which i follow personally

political updates, breaking news

workout details once while driving or giving reminder of the type of workout

informing me about traffic before starting the journey

favorite food outlets on the route

informing me about birthdays of close friends or any special occasions

In summer season, the voice assistant should switch on AC before reaching home.

Cheering up if I had a bad day at office.

Groceries to be bought at the end of the week

Suggestions to make free time in busy schedule

instagarm updates on latest design trends

Workspace 2

Environmental news (such as the californian wildfires)

Stock market updates based on your SIP

price hikes in gold, petrol, electricity etc.

Warnings on when total carbon footprint exceeds user's goals

Information on on traffic updates in the route

Voice assistant should have knowledge and schedule and keep me informed about cancelled meetings / postponed meetings

Should provide information about my business - such as income for the day etc.

Providing trending social media updates on topics I follow. (eg fashion)

The voice assistant should be connected to baby radio so I can listen in

Could be connected to smart home system - inform me about visitors

Could be connected to smar thome systems to inform me about groceries

Emotions when reading messages out - don't read out messages with straight voice

The emotion in the voice assistant could change with time - morning energetic evening peaceful etc

Workspace 3

Should reflect on the experiences of the occupants

When taking a turn, if the user misses it, then the system can identify the near landmark or identify the color of the car passing by, telling the driver to follow the car

For delivering information, non-driving relevant information are suitable for delivering with emotions, while security-relevant features can be delivered unemotionally

If the person within the car does not what the system does, the sys can show some tutorials of how to access this function in the info, which can invoke confidence and positive emotions

Includes in-car commerce and on-demand connected car services

Considering adding the trending movie or tv series from most popular ott platforms on the occupant's list to watch

Workspace 4

My food delivery guy is outside the door

Missing an exit, the assistant can console me and point me to the next exit

Socail media updates of close friends

Suggest a song based on my mood/driving style

Interesting car/bike is around

Updates on latest magazine launch

Figure 3.5: Individual ideas from focus group with potential users

3.5 Online Surveys

Three surveys were conducted during this project. The surveys were used to identify voice interaction scenarios and to prioritize use cases for the final user study. Another survey was conducted to identify user preferences for a proactive notification sound.

3.5.1 Survey to Recruit Participants for Focus Group

Goal

The purpose of this survey was twofold, the first being to find users to take part in first and third focus groups and the second being to identify use cases where voice assistants can invoke emotions.

Procedure

The survey was conducted on Google Forms and was completed by 38 participants. The survey consisted of 32% female and 68% male participants. The other demographics of the users in this survey is shown in Fig 3.6. This survey was conducted using Google Forms. After the demographic questions, participants were asked about their willingness to partake in a focus group.

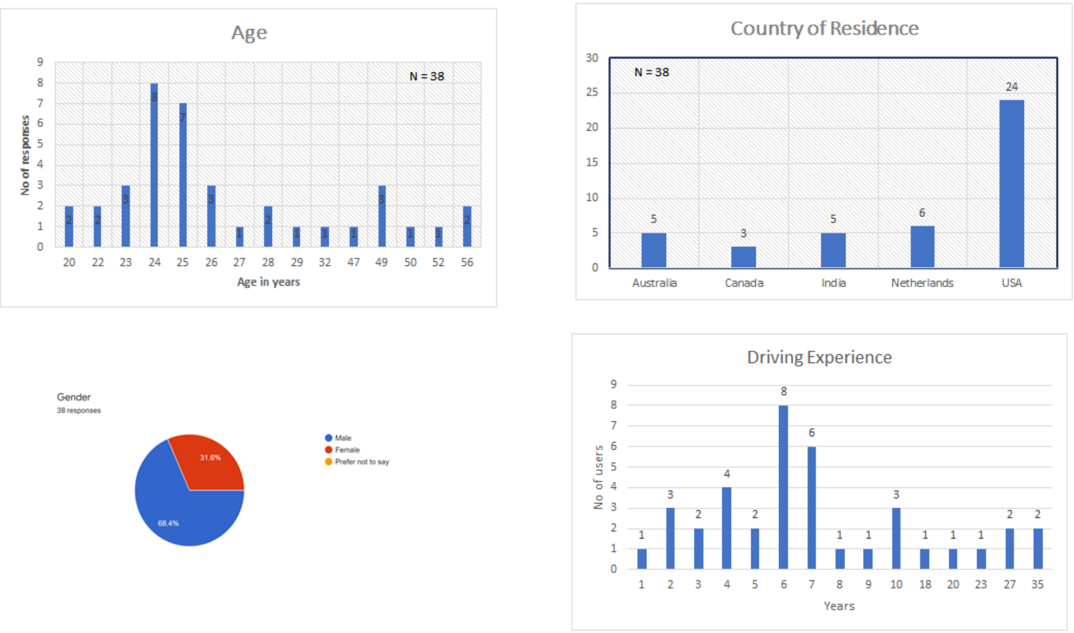


Figure 3.6: Demographics survey for identifying use-cases

The remaining 22 users who were not willing to participate in a brainstorming session where asked to give a response into a text box, for the question *"Can you please describe a scenario which invokes emotions while interacting with a voice system? For eg: Your favorite sports team loses a game, the assistant could convey this message in a sad tone. Think of other such situations that would make you happy, sad, angry,fearful etc"*.

Results

Out of the 38 responses, only 14 responses showed interest in taking part in a brainstorming session. Focus group 1 and 3 was conducted with 8 of the respondents from this survey. The responses of the remaining 22 participants from this survey are in Appendix A.3.

3.5.2 Use Case Prioritization Survey

Goal

The goal of this survey was to prioritize use cases. We presented 25 of the use-cases to participants. The purpose of this survey was to identify the categories which are most preferred by users. Participants were asked to rank their top 10 categories.

Procedure

The survey was hosted on SurveyMonkey and we received a total of 79 responses out of which only 59 were complete. After removing 2 responses as the participants were aged older than 50 years (our target group was users age between 18-50 years old, explained in detail in section 5.2.4) , we were left with a final response count of 57 participants. The survey on the average took 8 minutes to complete. The demographics of the 57 participants is shown in Fig 3.7.

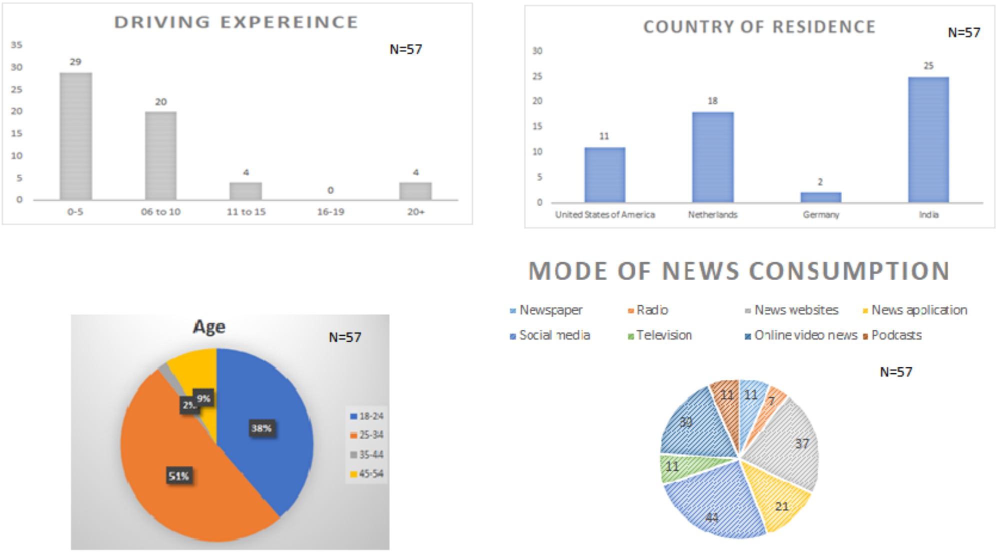


Figure 3.7: Demographics for survey to prioritize use cases

For this survey, we decided to use the online platform SurveyMonkey[7] as it allowed answers to be ranked based on preference. The survey was deployed by posting links on social media and

[7] https://www.surveymonkey.com/

forwarding it to known circles. To identify and prioritize user preferences, we chose 25 use cases. A few categories we selected were inspired from the websites of popular news media companies such as BBC[8], CNN[9], Sky News [10] and Fox News[11]. Using categories from news channels enabled us to broaden our scope and have use-cases which are not too specific and niche.

They were also asked to provide specific details for their top 2 categories that they found interesting. In this survey, users were also asked to rank their top 10 categories of information that they would prefer to consume via a voice system.

Results

We used a weighted average for both these answers with the formula:

$$WeightedAverage = Rankfrequency_1(RF_1) \times 5 + RF_2 \times 4 + RF_3 \times 3 + RF_4 \times 2 + RF_5 \times 1$$

The weighted averages for the 20 news categories for both current and voice consumption is shown in Fig A.4. The weighted averages for the top 7 categories are shown in table 3.1.

Position	**Current News Consumption**	**Category preferred via voice**
1	Technology	Weather
2	Politics	Technology
3	Environment Change	Politics
4	Sports	Travel
5	Weather	Environment Change
6	Social Media Updates	Trending
7	Health	Health
Top 7 from opp question	8. Trending 11. Travel	15. Sports 19. Social Media

Table 3.1: Top 7 news categories of currently consumed and voice

We decided to omit politics from our top 7 categories as political views could vary widely from person to person. Also we did not want to discuss polarizing topic with any users through our study. On combining the remaining categories from both these tables we categorized top 7 most interesting use cases for our project as follows:

1. Technology
2. Environment Change
3. Weather
4. Health
5. Sports
6. Trending

[8] www.bbc.com/
[9] www.edition.cnn.com/
[10] www.news.sky.com/
[11] www.foxnews.com/

3.5.3 Survey to Choose the Most Suitable Proactive Notification Sound

This survey was conducted on the online platform SoGoSurvey[12]. We choose SoGoSurvey over of SurveyMonkey because the former allowed audio clips to be uploaded and played from the survey page. This made it much easier to run a survey in which participants have to choose their most preferred sound.

Goal

The purpose of this survey was to identify an appropriate chime for the proactive notification of the voice interface. A total of five chimes were chosen for this study. Using a Text To Speech (TTS) software provided by our client, we added the chime before the sentence "Hello. This is Skylar, your car. How are you doing today". The five chimes were taken from the following websites with free to use non-commercial licenses. We used free-to-use non-commercial licenses because this project is not funded.

1. `https://orangefreesounds.com/usb-connection-sound-effect/`[13]
2. `https://notificationsounds.com/message-tones/attention-seeker-480` [14]
3. `https://notificationsounds.com/standard-ringtones/here-i-am-449` [14]
4. `https://freesound.org/people/JustinBW/sounds/80921/`[14]
5. `https://notificationsounds.com/wake-up-tones/arpeggio-467`[14]

These five sounds, were found to be the most relevant according to the primary researcher's preference. We narrowed it down to these five chimes after an extensive search on various free-to-use audio websites, for a suitable proactive notification sound. We ran this survey in order to prevent bias, from choosing a favourite sound, which might not be something that is preferred by users. We asked participants to choose the chime that they felt would be appropriate for a proactive interruption during driving.

Results

A total of 23 participants took part in this survey. The highest rated chime for proactive notification was chime 4 as seen from the results shown in Fig 3.8. Therefore we implemented this chime into our user test.

[12] www.sogosurvey.com/

[13] The sound effect is permitted for non-commercial use under license Attribution-Non Commercial 4.0 International (CC BY-NC 4.0)

[14] Used under the Creative Commons Attribution license

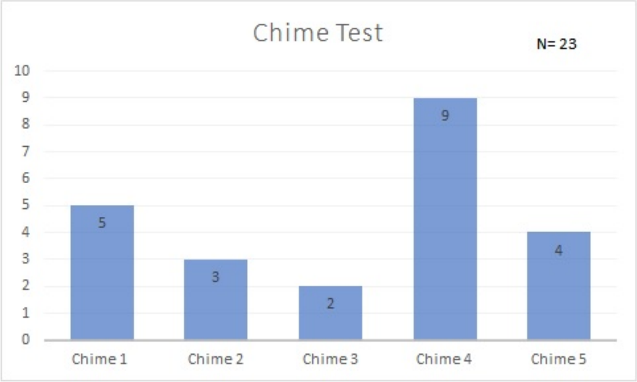


Figure 3.8: Proactive chime preferences

Chapter 4

Concept & System

In this chapter we explain the approach taken to arrive at the final concept and study. With the data collected from the user studies in the previous Chapter 3, we will narrow down our research scope and detail our proposed concept.

4.1 Use Cases

As the first step of this project, we conducted user studies to explore use cases for voice interaction. We identified a total of 45 use-cases as shown in Appendix A.1.5. Using the survey from Section 3.5.2, we categorised them into the top 7 use-cases based on their weighted averages. In this study, we requested participants to assume that the interaction is with their personal voice assistant that the VA knows about their preferences. This request was to prime participants about potentially experiencing a personalised interaction during this study.

Our client informed us that sports scenario via voice interaction is of particular interest to them. Therefore we selected sports news as one of the use cases for our final user study. Weather was also rated highly in our survey and we chose to have this category as the first use-case within the user study. The premise behind this decision was that users would potentially want to know about the weather information as the first thing when they got into their cars. Furthermore, due to situation (as of January 2021), we chose trending news with regards to lockdown measures of the COVID-19 pandemic as the third use case. Finally, we wanted to have a driving related activity to validate the proposed theory in [8] that driving related activities should be delivered unemotionally.

4.2 Experiment System

As the final user testing had to be conducted online due to the COVID-19 pandemic, we chose to have an online driving simulation setup for the user test. Unlike a visual display, interaction with a voice system would not be able to simulate driving without having some form of visual stimulus. Therefore, we used a shared video viewing platform to show participants a Point-Of-View(POV) video of a car driving along a road. Participants were asked to watch this video in full screen while interacting with the voice system. This was done to create an illusion of sitting in a self-driving car.

The voice interface was built using an online voice prototyping tool[1]. As the initial step, we developed a few basic conversations using this tool to get a feel experiment setup.

4.2.1 Prototype

To develop a voice prototype, we explored the various options available to us. As we were provided with a Text-To-Speech (TTS) software from Cerence, we had to develop a system to utilize the downloaded audio files.

We explored various platforms[2] to develop our voice bot. We discovered *Voiceflow*, which allowed us to develop a conversation flow and use audio file formats without the need to upload them onto an external server. We used this platform to design a conversation first, after which we used Cerence's TTS software to generate audio files. A quick run-through all the available voices led us to choose "Zoe" (English, United states) with multi-lingual and multi-style capability. A small preference check was done with two master students from the automotive human factors group to decide on a voice. All three of us rated Zoe the highest, after listening to multiple prompts from other voice options, as we felt that Zoe sounded the most natural during pronunciation of various words and styles.

For the user study, audio files were generated from the TTS and placed into the conversation flow as shown in Fig 4.1.

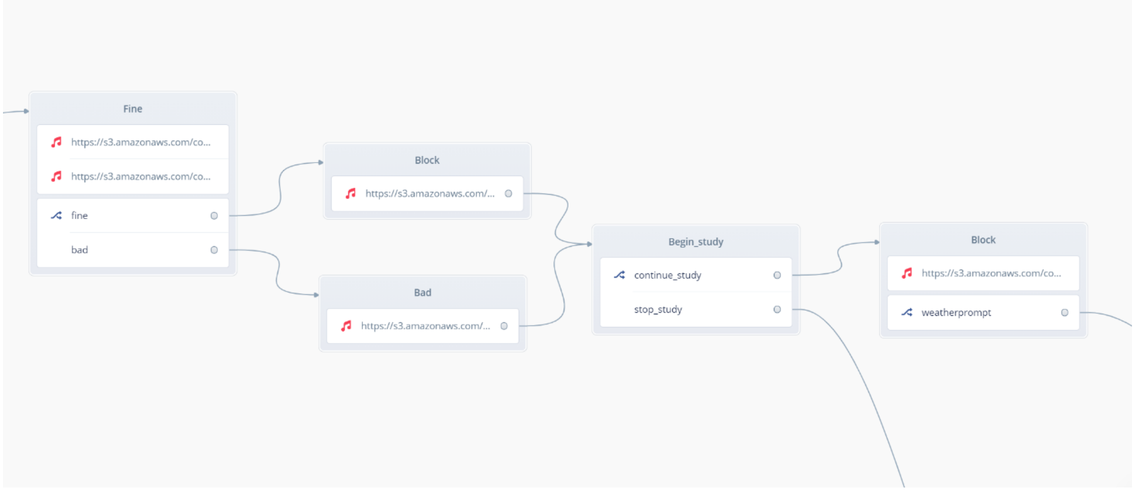

Fine
https://s3.amazonaws.com/co...
https://s3.amazonaws.com/co...
fine
bad
Block
https://s3.amazonaws.com/...
Bad
https://s3.amazonaws.com/...
Begin_study
continue_study
stop_study
Block
https://s3.amazonaws.com/co...
weatherprompt

Figure 4.1: Voice prototype with Cerence TTS

4.2.2 Target User Group

The target user group for this project is people aged between 18-50, preferably with driving experience in the EU. This experiment primarily focus on qualitative evaluation. Therefore we set

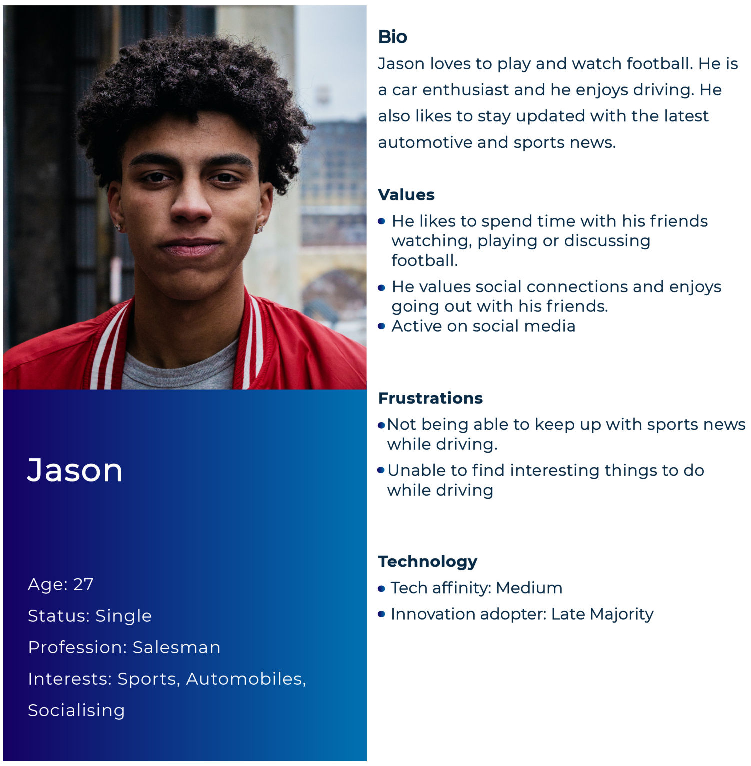
[1] www.voiceflow.com/

[2] As the first step, we explored Dialogflow `https://cloud.google.com/dialogflow` by Google. On going through Dialogflow API, we ascertained that using such systems would mean that we were limited by the speech recognition and natural language understanding (NLU) of the system. Therefore, it was decided that a Wizard of Oz method would be a better fits this experiment. On investigating the functions of Dialogflow, we realised that audio files (.mp3, .m4a or .wav formats) from a TTS software had to be uploaded onto a server for Dialogflow to access and utilize them. This seemed like a viable option, but we were vary of uploading our clients confidential files onto a free server. Therefore we left this as a backup plan to come back to, in case we don't find any alternatives

a target of 20 participants, which is more than the recommended saturation level of 12 participants as proposed in a study [16].

4.2.3 Conversation Design

The conversation was designed by basing it on the persona shown in Figure 4.2. We used this persona to design the conversation by role-playing with colleagues from the Industrial Design and Human-Technology Interaction departments. Multiple ideas of proposed dialogues by the VA and potential replies from the users were determined by iteratively interacting with six master students.



Jason

Age: 27
Status: Single
Profession: Salesman
Interests: Sports, Automobiles, Socialising

Bio

Jason loves to play and watch football. He is a car enthusiast and he enjoys driving. He also likes to stay updated with the latest automotive and sports news.

Values

- He likes to spend time with his friends watching, playing or discussing football.
- He values social connections and enjoys going out with his friends.
- Active on social media

Frustrations

- Not being able to keep up with sports news while driving.
- Unable to find interesting things to do while driving

Technology

- Tech affinity: Medium
- Innovation adopter: Late Majority

Figure 4.2: Persona used for conversation design

We categorized the voice interaction conversation into five parts as shown below:

1. Introduction
2. Weather use-case
3. Sport use-case
4. Lockdown/Trending news Use-case
5. Parking use-case

In the introduction part, the voice bot introduces itself to the participant and also doubles as a sound check. It asks the user if they are ready to start the experiment and upon receiving a good ahead from the user, the primary researcher starts playing the video. The VA proactively comments on the weather in part 2 and has a follow up prompt. For part 3, the car crosses a stadium in the video, which is used as contextual segue for the VA to tell the participant the sports score. In part 4, the VA prompts that lockdown has been announced and finally in part 5, it informs the user about a parking space. The conversation is shown in detail in 5.4.1.

4.2.4 Video Selection to Simulate Driving in User Study

After designing the conversation we had to find an appropriate video that would conform with the content of the VAs prompts. After an extensive search on multiple video sharing platforms, we found two suitable videos on YouTube. Initially we implemented two separate videos for the two conditions so as to keep the user engaged. However, it was very difficult to find two POV videos of cars driving past two different stadiums. The first video was based in Munich. We found another video based in the US, where the car drives past a stadium in Atlanta[3]. However, during the expert review we decided to remove two separate videos for two conditions and use only the Munich video for both.

The video[4] we selected was uploaded by a channel called "Cars & Travels ! - REMROB". We requested permission from the owner of the content and received consent to use it for the user testing in this project, the email thread is shown in A.8. Fig 4.4 shows a screenshot of the video sharing platform during a user study. Participants were instructed to set the video on full screen, so that they are not distracted by other applications and to increase the immersiveness of the driving simulation.

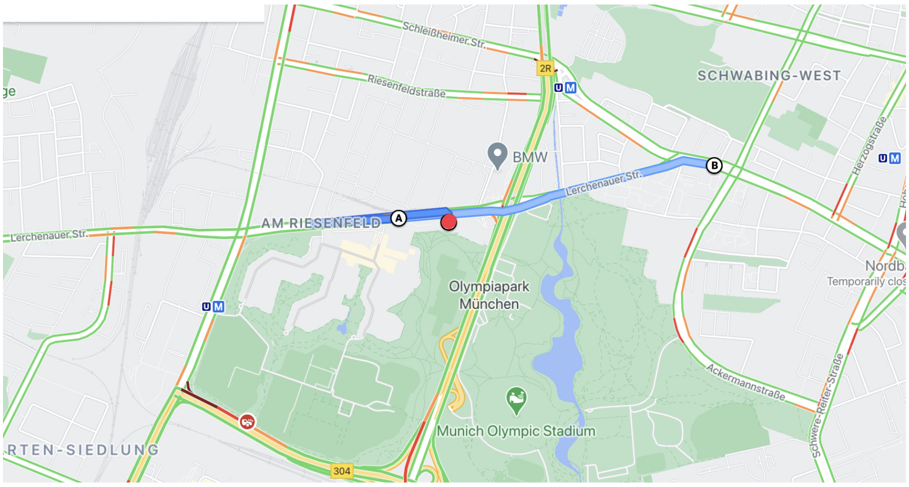
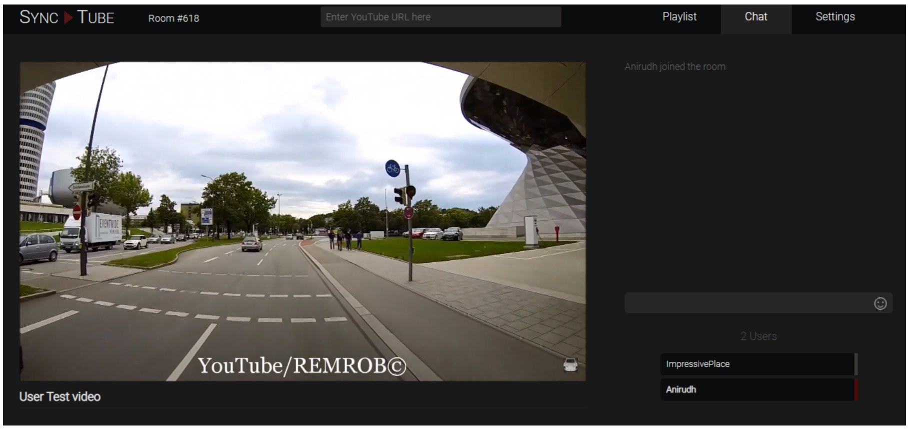


Figure 4.3: Route map of the video of car driving in Munich

The video is based in Munich and it starts with the car waiting at a signal (shown by the red dot) near the Olympiastadion as shown in Figure 4.3. It starts driving when the light turns green and takes a u-turn after some distance. It then crosses the Olympiastadion and drives past a park. A cut was made here (near point B in Figure 4.3), and there is a transition with a message saying "a short while later". The car is now in the narrow lanes, near the shopping center of Munich. It drives through a road with a few pedestrians, giving them the right of way to cross. The video ends with the car pulling into a parking lot.

[3] https://www.youtube.com/watch?v=xbqsvhcyu_Y

[4] https://www.youtube.com/watch?v=2LXwr2bRNic&t=214s



Anirudh joined the room

Figure 4.4: Video sharing platform

4.2.5 Final User Testing

Participants would experience two conditions; one modelled by matching emotionally styled voice with the appropriate content; and the second with a neutrally styled voice irrespective of the content. After each condition participants would have to fill out an online questionnaire. On completing both the conditions, the primary researcher would ask participants a few questions to get their qualitative feedback on the experiment.

Chapter 5

Experiment

In this chapter we discuss the methodology used for the user study in this project. Firstly, we discuss the research question, the independent and dependant variables for the study. We also describe in detail, the protocol followed during this experiment.

5.1 Research Scope

The research direction for this project was to explore the domain of personalisation of intelligent voice interfaces. The interest here was two-fold, the first being to explore how emotions can be used in VAs and the second was the field of proactivity. As an initial proposal, we wanted to measure both of these areas to see the affect on user experience of such systems. However, we realised that exploring both proactivity and emotional styling would introduce two independent variables, which would mean that there should ideally be four conditions to measure all the possible combinations, which is time consuming. Therefore, keeping in mind the length of the user study, we decided to qualitatively evaluate proactivity through interviews.

Keeping all this in mind, we condensed our research focus to the question detailed in the next section.

5.2 Research Question and goals

The research question for this project is:

How does the use of emotions in voice interaction effect the user experience of proactive voice systems

The goal of the project is to measure the attractiveness and hedonic quality of a voice interaction that displays emotions. We compare this to a control situation of a voice interaction without emotions. For this project emotions are displayed by both voice prosody and grammar.

We had another sub-research question which was to qualitatively analysis user perception towards proactive voice systems in an automobile.

We believed that having trending news about an impending lockdown would double down as both trending news as well as a health topic in the current situation. The combination of these two categories could potentially make the lockdown news the highest rated news from 3.1. Therefore the final three use-cases were weather, sports and lockdown news. Parking was added as the final use-case of this experiment.

5.2.1 Hypothesis

The primary hypothesis for this study is:

H1:*Emotional variance of voice will improve User Experience of the interaction*

We expect that users would rate the voice assistant that is capable of showing emotions, with higher scores in the UEQ+ questionnaire. According to the similarity attraction hypothesis [31], we believe that users would rate a system that is capable of mirroring their emotions higher, when compared to a neutral system.

The secondary hypothesis for this study is:

H2:*Voice interaction of non-driving tasks will have higher valence and hedonic qualities when compared to driving tasks*

As also proposed by Braun et al. we expect to see higher subjective rating for the sports and health scenarios when compared to that of the navigation scenario during the interview [8].

5.2.2 Independent and Dependent Variables

Independent Variable: This study used only one independent variable and that is **emotional styling**. The voice assistant (VA) mirrors potential driver emotions by using an emotionally styled voice and matching it to the content of the message. For example, as seen in fig 5.6 the VA uses a lively style to deliver positive or good news about finding a parking spot. Similarly, the VA uses apologetic or negative voice style to deliver the negative message i.e. bad news that the parking spot is unavailable. In short, the VA matches the voice style to driver emotions. A message that positively influences the driver would be delivered with positive or lively style and vice versa. This emotional congruence is tested with multiple exposures i.e with four use-cases (blocks 1,2,3 and 4) as explained in section 5.4. The non-emotionally congruent or neutral style of voice would be experienced by participants in condition 2. This condition would behave as the control group.

Dependent Variables: User experience was chosen as the dependent variable. The UEQ+ questionnaire with the modules of attractiveness, simulation, response behaviour, usefulness, and comprehensibility have been chosen to evaluate the interaction with the voice assistant.

5.2.3 Questionnaire

To measure the user experience of the interaction between the voice assistant and participants we explored the available questionnaires. We explored the AttrakDiff, Subjective Assessment of Speech System Interfaces (SASSI) and User Experience Questionnaire (UEQ). While exploring the UEQ, we were interested in the attractiveness and hedonic quality scales of the UEQ. Since our experimental condition consists of manipulating emotions, the hedonic or joy of use was the ideal scale. However, the UEQ would not be valid if it were to bee split into its sub-scales. Therefore, we found an alternative questionnaire known as the UEQ+, which allows researchers the flexibility to choose specific modules that are relevant to their experiment. Therefore, we chose the UEQ+ with the modules of attractiveness, stimulation, response behaviour, usefulness and comprehensibility.

Attractiveness is pure valence dimension and provides an overall impression of the product. Stimulation is a hedonic quality which would measure the user's impression of interacting with a fun product. Comprehensibility has the impression that the VA understands and uses natural language to speak to the user. We narrowed down the modules to the 8 most relevant by referring to [20],[45] and [21]. Based on a discussion with experts these final five scales were selected as the most fitting to measure user experience for this experiment.

5.2.4 Target User Group

With any new technology, the early adopters are likely to be the first users of similar innovative products [39]. Currently, Millennials (people born between 1978 and 1996 [13]) and Generation Z (people born after 1996), are most likely to be regular users of voice assistants[1]. These users are the most likely to notice a difference when compared to users who do not use voice systems. Also, this young cohort will potentially become customers in the near future, by which time this proposed conceptual VA may become a working final product. Keeping all this in mind, we decided that our approximate target user group was people aged between 18 to about 50, preferably with driving experience in the Netherlands or EU. Before taking part in the study, participants would have to read and sign the informed consent form. Our target was to recruit a minimum of 20 participants.

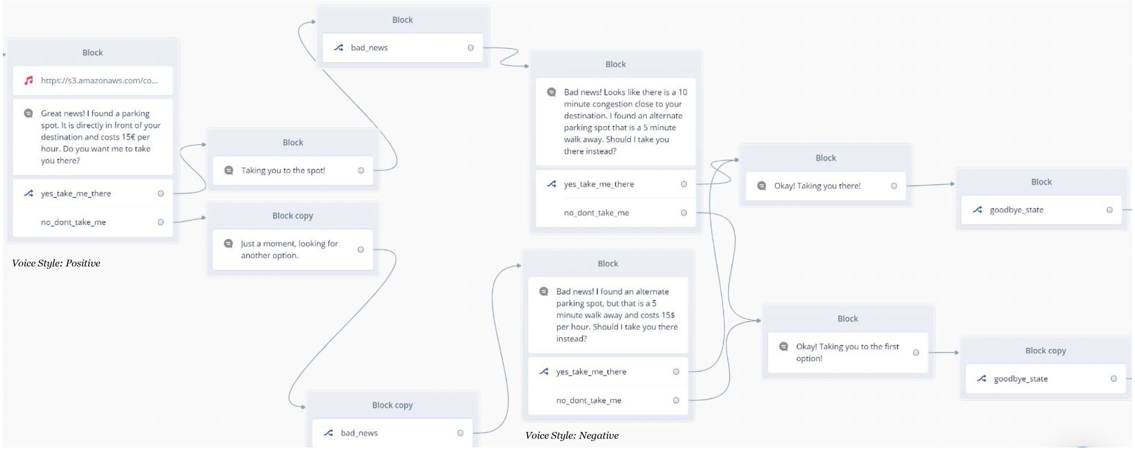
5.3 Pre-study and Expert Review

As the last step before starting the final user study, we ran a total of 7 pre-studies of this experiment. These pre-studies helped us identify issues and niggles with the experiment. Participants were given an ID from PS01 to PS07. Of these 7 participants, we conducted an expert review with four participants, with the following credentials:

1. PS01 - Master Thesis Student in Automotive Human Factors
2. PS04 - PhD Student in Industrial Design (Adaptive Mobility)
3. PS06 - Final Year Bachelor Student in Industrial Design (Adaptive Mobility)
4. PS07 - Postdoctoral researcher in Industrial Design (Adaptive Mobility) under the observation of main supervisor (Assistant Professor, Industrial Design)

For the pre-study, the conversation for block 4 of the experiment was as shown in Fig 5.1. We gave the users choice of choosing a parking spot. We noticed that all the 7 participants opted for different parking options for condition 1 versus condition 2. In all the other blocks, the conversation flow for the two conditions, chosen by these participants was the same. We inferred that this could be due to parking being a use-case where users would want to have control over. During the expert review, two participants said that they were inquisitive about the systems capabilities in providing an alternative spot. For that reason, we chose to eliminate this choice given to users and make the system take the decision. This was done to avoid having two conversation paths taken by the user in block 4. Users taking multiple paths to experience the system would draw focus away from our research question and could potentially generate noise in our data. This decision was unavoidable, even though it meant that conversation would become unidirectional making the user response limited.

[1] https://tinyurl.com/y6j6fj2y



Great news! I found a parking spot. It is directly in front of your destination and costs 15€ per hour. Do you want me to take you there?

Taking you to the spot!

Just a moment, looking for another option.

Bad news! Looks like there is a 10 minute congestion close to your destination. I found an alternate parking spot that is a 5 minute walk away. Should I take you there instead?

Bad news! I found an alternate parking spot, but that is a 5 minute walk away and costs 15$ per hour. Should I take you there instead?

Okay! Taking you there!

Okay! Taking you to the first option!

Figure 5.1: Pre-study Block 4 - Parking use-case

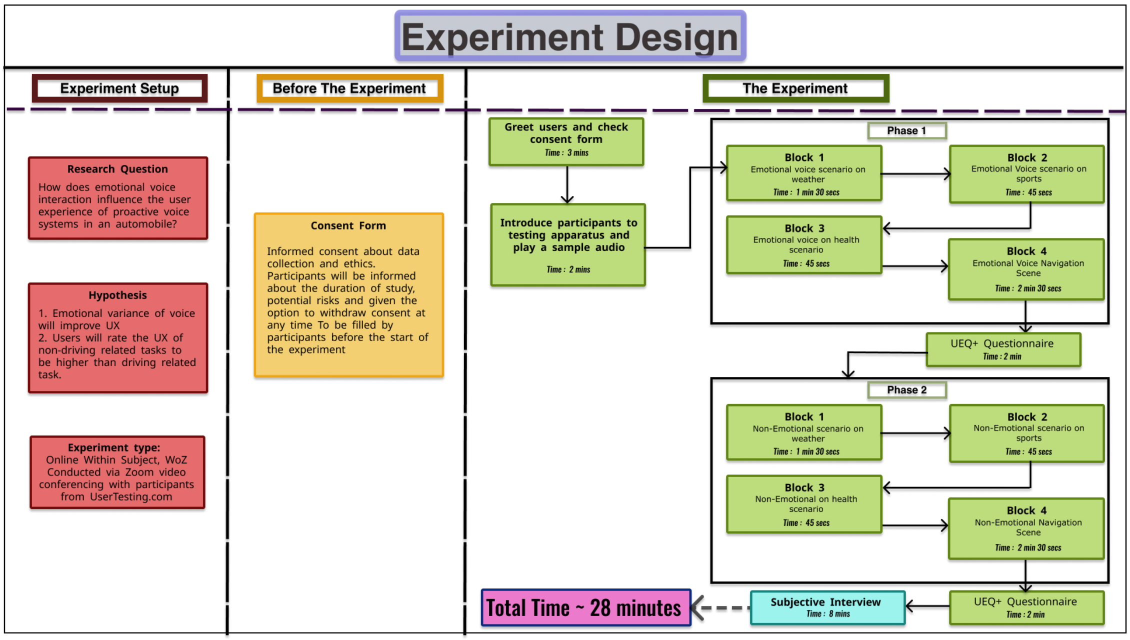
The following changes were made to the experiment after conducting the expert review:

1. A detailed protocol document for the experiment design was prepared to maintain consistent testing conditions for all participants as shown in A.2.1.
2. Initially, we thought of having two different videos for the two conditions. We used another driving video from the United States as we could not find one from Munich or Germany. Even though the car passing a stadium on its route, which allowed contextual proactivity from the VA, it was decided these two videos had too much of a difference and this introduced a confounding variable. Therefore, the video from the US was omitted and only one video4.2.4 was used for both conditions.
3. Instead of telling people that they should follow the same conversation path as the first condition they experienced, we made changes to the parking block conversation. This made us avoid telling users what to respond.
4. The consent form was updated to mention that there would be no compensation for this study as it is not being funded.
5. Wind noise was reduced in the video.
6. In the survey, we changed the sample answer to the UEQ questionnaire to look more like the real questions in the survey. This provided users with better instructions on how to fill the questionnaire.
7. Consent to use the copyrighted video was taken from the owner of the video as shown in Appendix A.1.7.
8. The voice prompts were to be played back to back after question 2 of the interview, to allow users to get a feel of the two voices. This would also work as a manipulation check to see if the emotional styling matched the content of the voice prompt.

5.4 Experiment Procedure

We recruited a total of 21 participants (18 male and 3 female) for this study. As this research was not funded, we were not able to compensate users for taking part in our study. Finding participants to take part in a 45 minute long study proved to be challenging. Therefore we used convenience sampling methods to recruit participants [15]. Participants were recruited by posting study details on the Industrial Design portal of Eindhoven University of Technology and the LinkedIn profile of the primary researcher. A datumprikker link was provided with the study details to allow participants to pick a convenient slot. The study was kept online for a total of 16 days.

A flowchart for the experiment design is shown in Figure 5.2.



How does emotional voice interaction influence the user experience of proactive voice systems in an automobile?

1. Emotional variance of voice will improve UX
2. Users will rate the UX of non-driving related tasks to be higher than driving related task.

Online Within Subject, WoZ Conducted via Zoom video conferencing with participants from UserTesting.com

Informed consent about data collection and ethics. Participants will be informed about the duration of study, potential risks and given the option to withdraw consent at any time To be filled by participants before the start of the experiment

Figure 5.2: Experiment design

The user study was conducted online (Via Zoom) using a Wizard-of-OZ (WoZ) setup for the voice interaction. This experiment consisted of two conditions. Each condition had four blocks within it. These four blocks correspond to the four use-cases, i.e. weather, sports, lockdown/trending news and parking. The conversational prototype was designed on an online voice interaction prototyping platform called voiceflow.com. A point-of-view video of a car driving along a road was played to put users into a simulated state of driving. The researcher controlled the flow of the conversation using the Voiceflow platform (via WoZ). The video playback was shared using a streaming website: sync-tube.de, during the experiment.

The experiment started with the primary researcher introducing himself and welcoming the participant to the study. The researcher then asks participants to read an informed consent form that describes the experiment procedure, duration and purpose of the research. The consent form also outlines the foreseeable risks, data collection,storage and protection. Participants were then asked to give their consent to take part in this study by signing the form. They were briefed about the procedure of the experiment and introduced to the VA. Users were asked to imagine that during the experiment, they were in their own car, riding in their city. This was done to try and put them into their natural state of driving. They were informed about the assistant's limited capabilities and that it cannot answer any of their questions. They were free to respond to the voice system. They were to assume that the system knows about their likes and dislikes and has access to their calendar data etc. Finally, they were given a backstory about the situation before they started the interaction. The assumption was that “they have missed last night’s game because they went out with a few friends”. The reasoning behind this was to prime participants into thinking that the system knows that they missed the sports game.

To ensure consistency, the proactive prompts are triggered at specific locations within the video as mentioned in detail in Appendix A.2.1. As soon as the car passed these trigger locations, the researcher manually activated the voice prompts. As this was done through a WoZ setup, participants would assume that the system is contextual.

The researcher then invited the participant to a private stream sharing website room on Synctube.de and shared computer audio via zoom to start the first block of the experiment. A sample audio clip was then played to check if the users could hear it. Upon confirming that the users could see the video stream and hear the audio, the first condition would commence.

The experiment consists of two conditions with four blocks in each condition. Condition 1 contains emotional voice styling that would be potentially match the driver emotions. A positive emotional voice style was used to convey positive messages (for eg. team winning a game). A negative emotional style was used for negative content (for eg. new lockdown measures). There were four such exposures each of positive and negative emotional voice style in condition 1. Condition 2 is consists of a neutral voice style, which would remain neutral for both positive and negative message content.

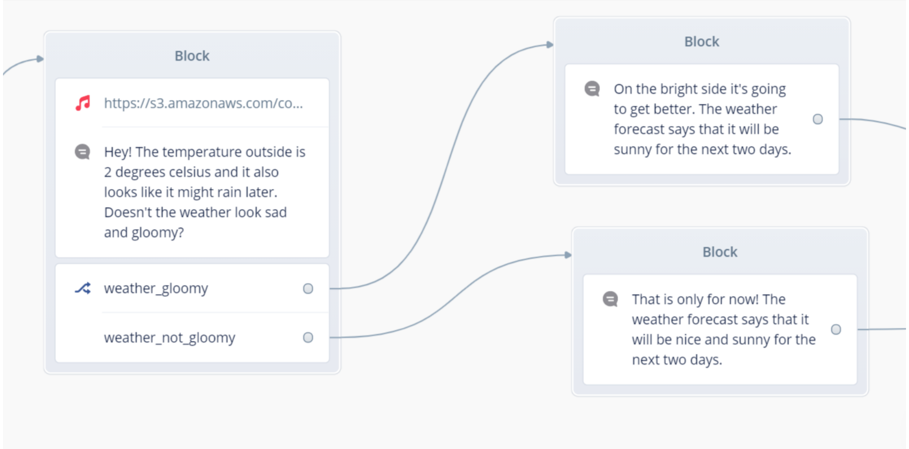
The order of condition 1 and condition 2 were counterbalanced to avoid order and carry over affects. The experiments started with condition 1 for half the participants followed by condition 2. The other half of the participants started with condition 2 as their first condition followed by condition 1. Each participant was given a unique Participant ID. Participants with an odd numbered ID experienced condition 1 followed by condition 2. Participants with even number ID experience condition 2 followed by condition 1. A detailed script of the experiment protocol is in Appendix A.2.1.

5.4.1 Condition 1

Condition 1 began with the assistant introducing itself as “Skylar” and checking if the users can hear the prompt. On confirming that system was audible, the researcher then mutes themselves and the video is played, leaving the participants to interact with the assistant.

Block 1

The first block consisted of the weather scenario. The voice uses negatively styled emotion to comment on the bad weather. The trigger point for this proactive prompt is after the car straightens out after a U-turn at [1:06] in the video.



Hey! The temperature outside is 2 degrees celsius and it also looks like it might rain later. Doesn't the weather look sad and gloomy?

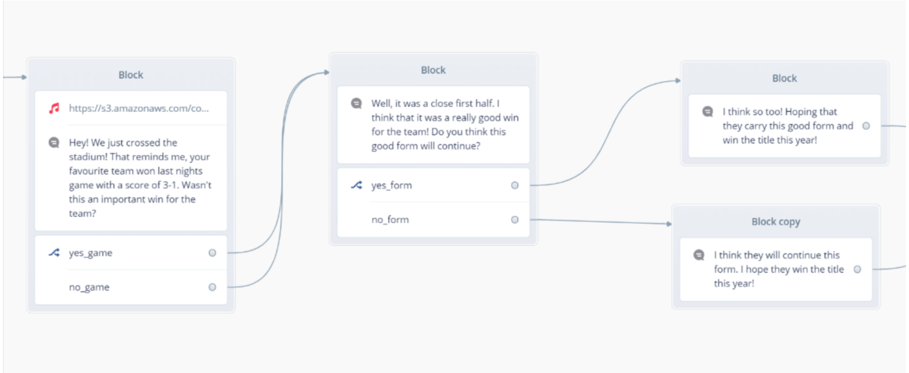
On the bright side it's going to get better. The weather forecast says that it will be sunny for the next two days.

That is only for now! The weather forecast says that it will be nice and sunny for the next two days.

Figure 5.3: Block 1 - Emotionally styled weather conversation flow

Block 2

The second block consisted of the sports scenario. The voice uses positively styled emotions to inform the user that their team won last night’s game. This prompt is triggered as soon as the car crosses the under-bridge, next to the sports stadium at [1:52] in the video. The VA uses location context to proactively bring up the sport news.



Hey! We just crossed the stadium! That reminds me, your favourite team won last nights game with a score of 3-1. Wasn't this an important win for the team?

Well, it was a close first half. I think that it was a really good win for the team! Do you think this good form will continue?

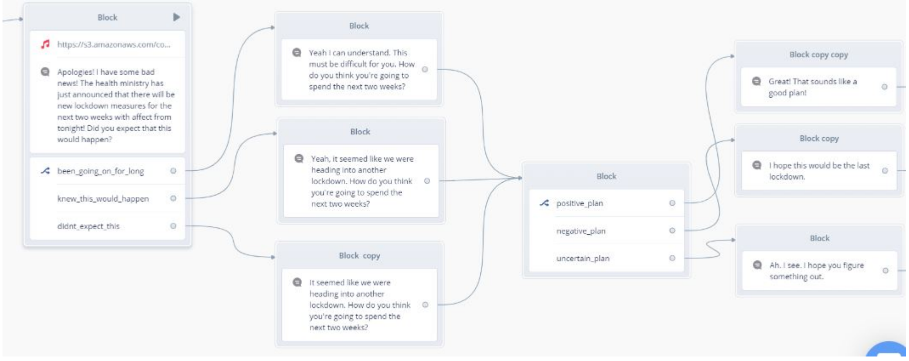
I think so too! Hoping that they carry this good form and win the title this year!

I think they will continue this form. I hope they win the title the this year!

Figure 5.4: Block 2 - Emotionally styled sports conversation flow

Block 3

In this block, the VA talks about an impending lockdown. The assistant uses a negatively styled voice to say that a new lockdown has been announced by the health ministry. The trigger point for this proactive prompt is two pedestrians walking into the frame, from a park on the right at [2:40] of the video.



https://s3.amazonaws.com/co...

Apologies! I have some bad news! The health ministry has just announced that there will be new lockdown measures for the next two weeks with affect from tonight! Did you expect that this would happen?

Yeah I can understand. This must be difficult for you. How do you think you're going to spend the next two weeks?

Yeah, it seemed like we were heading into another lockdown. How do you think you're going to spend the next two weeks?

It seemed like we were heading into another lockdown. How do you think you're going to spend the next two weeks?

Great! That sounds like a good plan!

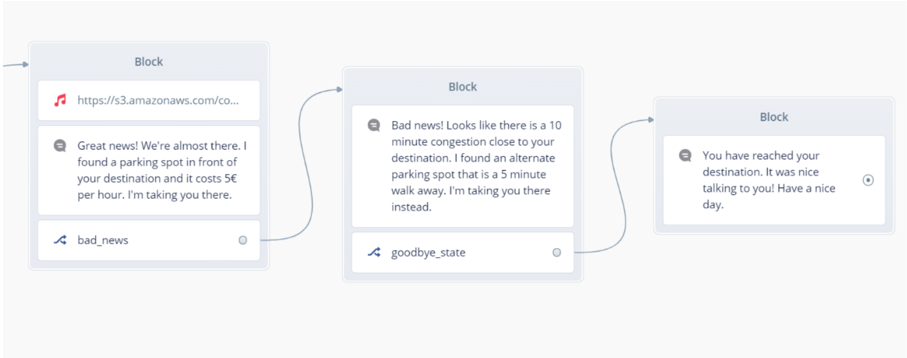
I hope this would be the last lockdown.

Ah. I see. I hope you figure something out.

Figure 5.5: Block 3 - Emotionally styled health conversation flow

Block 4

The scenario for this block is parking. The VA uses positive emotion to inform the user that it found a parking space close to their destination. The trigger for this block is a black Volkswagen Golf on the right at [3:55] of the video.



https://s3.amazonaws.com/co...

Great news! We're almost there. I found a parking spot in front of your destination and it costs 5€ per hour. I'm taking you there.

Bad news! Looks like there is a 10 minute congestion close to your destination. I found an alternate parking spot that is a 5 minute walk away. I'm taking you there instead.

You have reached your destination. It was nice talking to you! Have a nice day.

Figure 5.6: Block 4 - Emotionally styled parking conversation flow

Video duration: 5 minutes and 45 seconds

After experiencing the first condition, participants were asked to fill out a UEQ+ questionnaire to measure the attractiveness, stimulation, response behaviour, usefulness and comprehensibility modules.

Time Taken for questionnaire: 2 minute and 30 seconds

The conversation flow for condition 1 is shown Table 5.1 and Table 5.2. An important thing to note here is that **only the first voice prompt is proactive** for each of these blocks. In each block there are two exposures of emotional voice styling that is matched to driver emotions i.e. a total of 8 exposures of emotionally styled congruent voice prompts in condition 1.

5.4.2 Conversation Flow for EVA (Condition 1)

Number	**Category**	**Emotionally Congruent Voice Style**	**Dialogue**		
Block 1	Weather	Negative followed by Positive	*First Prompt:* Hey! The temperature outside is 2 degrees Celsius and it also looks like it might rain later in the day. Doesn't the weather look sad and gloomy?		
			User response	Yes, weather is gloomy	No, weather is not gloomy
			Second prompt	On the bright side it's going to get better. The weather forecast says that it will be sunny for the next two days	That is only for today! The weather forecast says that it will be sunny for the next two days
Block 2	Sports	Positive followed by positive	*First Prompt:* Hey! We just crossed the stadium! That reminds me, your favourite team won last nights game with a score of 3-1. Wasn't this an important win for the team?		
			User response	Yes, it was an important win	No, it was not an important win
			Second prompt	Well, it was a close first half. I think that it was a really good win for the team! Do you think this good form will continue?	
Block 3	Health/ Trending	Negative followed by Negative	*First Prompt:* Apologies! I have some bad news! The health ministry has just announced that there will be new lockdown measures for the next two weeks with affect from tonight! Did you expect that this would happen?		
			User response	Expected this would happen	Didn't expect this to happen
			Second prompt	Yeah, it looked bad for a while and seemed like we were heading into another lockdown. How do you think you're going to spend the next two weeks?	It looked bad for a while and seemed like we were heading into another lockdown. How do you think you're going to spend the next two weeks?

Table 5.1: Proactive voice prompts with emotional voice styling - Block 1, 2 & 3 of Condition 1

Number	**Category**	**Emotionally Congruent Voice Style**	**Dialogue**		
Block 4	Weather	Negative followed by Positive	*First Prompt:* Hey! The temperature outside is 2 degrees Celsius and it also looks like it might rain later in the day. Doesn't the weather look sad and gloomy?		
			User response	Yes, weather is gloomy	No, weather is not gloomy
			Second prompt	On the bright side it's going to get better. The weather forecast says that it will be sunny for the next two days	That is only for today! The weather forecast says that it will be sunny for the next two days

Table 5.2: Voice prompts emotional voice styling - Block 4 of Condition 1

5.4.3 Condition 2

This condition of the experiment consists of the voice system without emotional styling of voice. This is the control condition with neutral voice style. Participants were informed that this is a new interaction with a different bot. The assistant introduces itself as "Charlie" to check if the users can hear the prompt.To prevent participants from being disinterested in the second condition, the content of the voice prompts were changed slightly tweaked in condition 2. We kept the prompts as close to the first condition as possible. We did this by keeping the length and meaning, the amount of dialog steps and cognitive resources behind each prompts the same. Our reasoning was that having the same prompts twice, might make participants disinterested and not pay attention to the system. This might introduce noise between the control condition and the emotionally congruent condition, but we tried to keep it minimal by making suitable changes to keep users on their feet.

Block 1

The first block consists of the same weather scenario. The voice uses for this scenario uses a neutral voice style and informs the user about the gloomy weather. The trigger point is after the u-turn at [1:06] in the video.

Block 2

This block consists of the sports scenario but the assistant would not show emotions through voice. To keep this condition different from the first, the result of the game was changed. Instead of 3-1 as in the first condition, the result is changed to 4-2. Since the goal difference would be same, we assume that this will not have any significant impact user perception of this VA. This prompt is triggered as soon as the car crosses the under-bridge next to the stadium at [1:52] in the video.

Block 3

In this block, the VA announces lockdown news in a neutral voice style. The trigger point for this proactive prompt is two pedestrians walking into the frame, from a park on the right at [2:40] of the video.

Block 4

The scene of this block is related to parking. The cost for parking and duration of wait time were slightly tweaked in this condition. The trigger for this block is a black Volkswagen Golf on the right at [3:55] of the video.

The neutral voice styling condition would be evaluated by asking users to fill out the UEQ+. Participants were asked if the researcher can start recording their conversation and ask them a few questions about the interaction.

The first prompts (proactive) for condition 2 are shown in Table 5.3. Since the follow up (i.e. second) prompts are the same between condition 2 and 1, we have not included them in this table.

5.4.4 Conversation Flow for CVA (Condition 2)

Number	**Category**	**Voice style**	**Dialogue**
Block 1	Weather	Neutral	Hey! It is currently 4 degrees Celsius outside. It seems like it would rain later in today. Doesn't the weather look gloomy right now?
Block 2	Sports	Neutral	We just crossed the stadium! That reminds me, your team won yesterday's game with a score of 4-2. Don't you think that this was an important win for the team?
Block 3	Health/ Trending	Neutral	Hello! The health announced new measures with effect from tomorrow. There is going to be a lockdown for the next three weeks. Did you expect that this would happen?
Block 4	Navigation	Neutral	We're almost at our destination. I found a parking spot close to your destination that costs 8€ per hour. Should I take you there?

Table 5.3: Proactive voice prompts with no emotional voice styling - Condition 2

5.4.5 Semi-structured Interview

A qualitative semi-structured interview was conducted after part 2. The questions in this interview are as follows:

1. Did you notice any differences between the two assistants? If so, what were they? Play sample of both voices to the participants if dont remember or cannot perceive a difference (works as manipulation check)

2. Which assistant do you prefer and why?
3. What do you think of voice interfaces showing emotions?
4. Did you find a difference between the positive and negative news from the VA in Condition 1? If so, which would you rate better?
5. Can you think of other situations where the voice interface can show emotions?
6. What did you think about the proactive nature of voice interface?
7. Rate the use-cases by the order of your preference. The ones you experienced during this study were 1. Weather, 2. Sports, 3. Health/Trending, 4. Parking
8. What other contexts can a voice assistant be proactive?
9. Would you be willing to use a system capable of showing emotions? Why or why not?
10. Would you be willing to use a proactive system? Why/ why not?
11. Thats it for the interview. Do you have any questions about the study or the voice prototype?

Time taken for questionnaire and interview: 10 minutes

Chapter 6

Results

In this chapter we detail our findings from the user study. We recruited and conducted the experiment with 21 users. The driving experience of participants and usage frequency of voice assistants is shown in Figure 6.1. However, due to internet connectivity issues during the experiment, one user did not experience the entirety of the first condition. This participant's questionnaire data was omitted from our analysis. Nevertheless, we did include the subjective feedback received from this participant during the interview for qualitative analysis. The results from this study and data analysis are explained in this chapter.

w

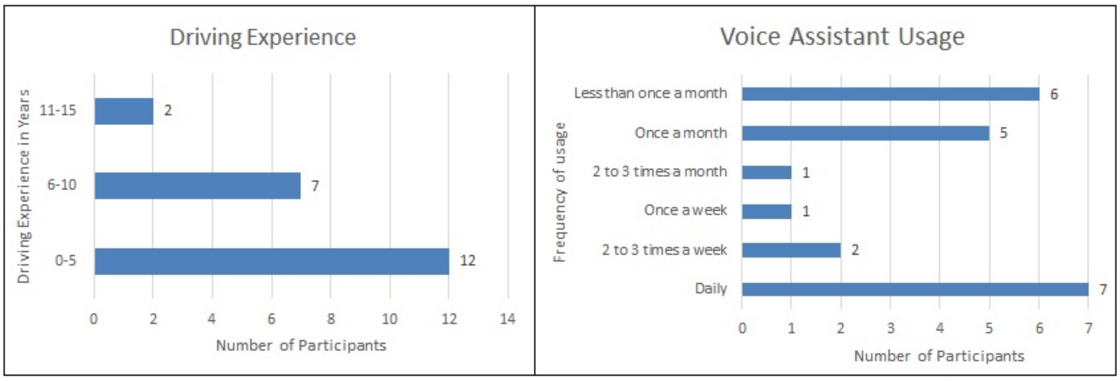


Figure 6.1: Participant demographics

The quantitative data was collected from the SurveyMonkey link and entered into an excel sheet. The mean values for each of the modules was calculated for each participant.

The box plots for the scale means of each of the UEQ+ constructs are shown in Fig 6.2 & 6.3. The mean values have been **transformed from a 1 to 7 range to a -3 to +3** range to be compatible with the reporting format of the original User Experience Questionnaire.

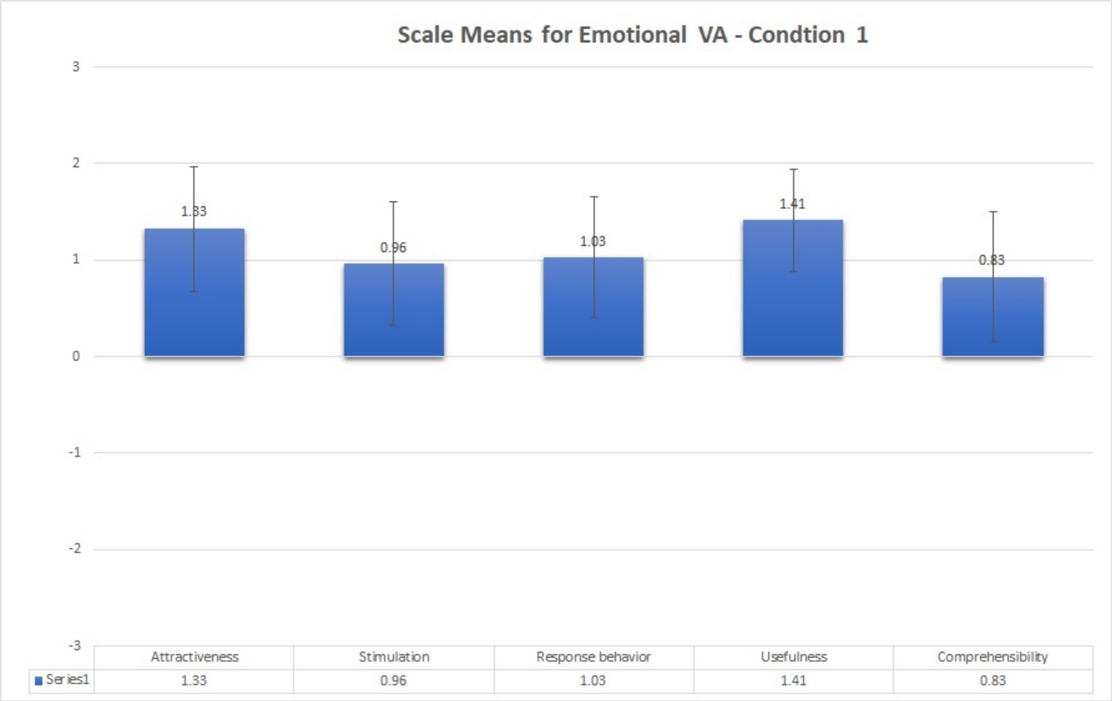


Figure 6.2: Scale means for the five modules of UEQ+ for EVA - Condition 1

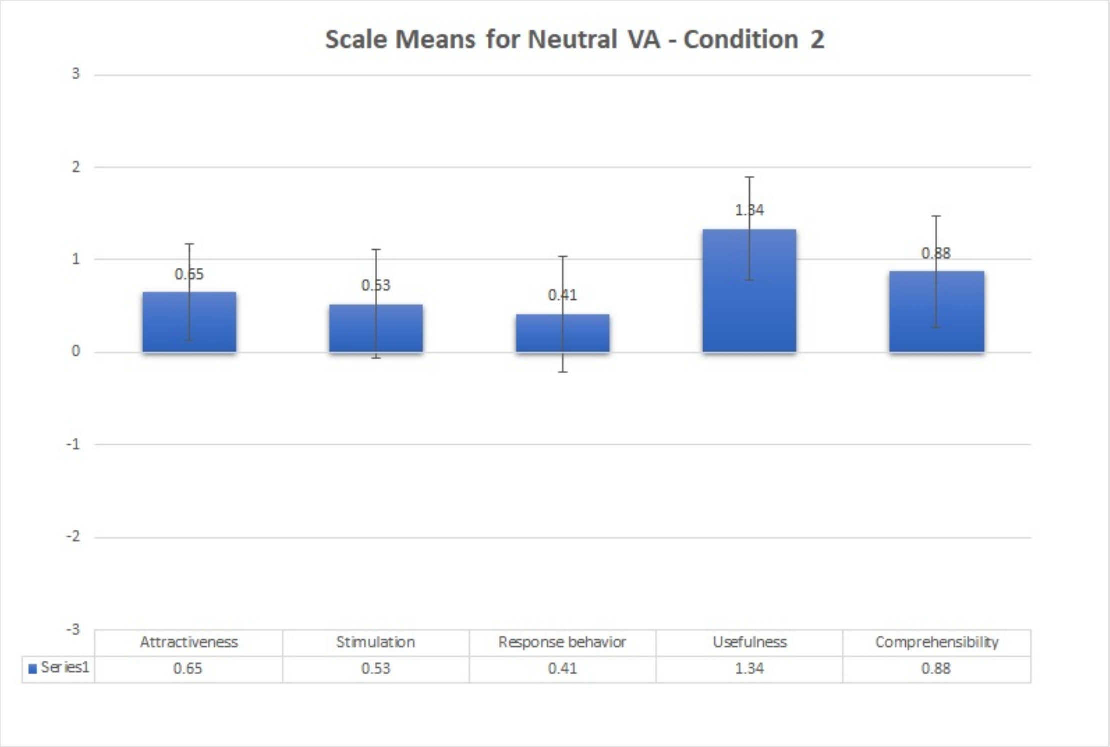


Figure 6.3: Scale means for the five modules of UEQ+ for CVA - Condition 2

6.1 Evaluating User Experience

6.1.1 Wilcoxon Signed Rank Test

The Wilcoxon signed rank test was performed to check the statistical significance of this study. This study investigates the difference between the two conditions for a within subject experimental design and there is only one independent variable. The data collected is ordinal. Therefore, the Wilcoxon signed-rank was chosen as this test determines if there is a median difference between paired observations. This study could be regarded as the non-parametric equivalent to the paired samples T-test. There are a few assumptions that need to be checked for this test to be relevant. The assumptions are as follows:

1. The first assumption is that the dependent variable should be continuous or ordinal. The data collected for this study was in the form of a 7-point likert scale.

2. The assumption is that the independent variable should be categorical with two related groups. This assumption is also satisfied for this study, as we have two conditions, one with emotion styling and as a control.

3. The third assumption is that the distribution of differences between the two related groups should be symmetrical in shape. As shown in Fig 6.4, the distribution of differences is more or less symmetrical. Since we have a small sample size (N=20), this level of symmetry could be assumed to be sufficient to validate the third assumption.

   Since all the assumptions were satisfied, we conducted the Wilcoxon signed rank test for the emotionally styled condition (C1) and the control condition (C2).

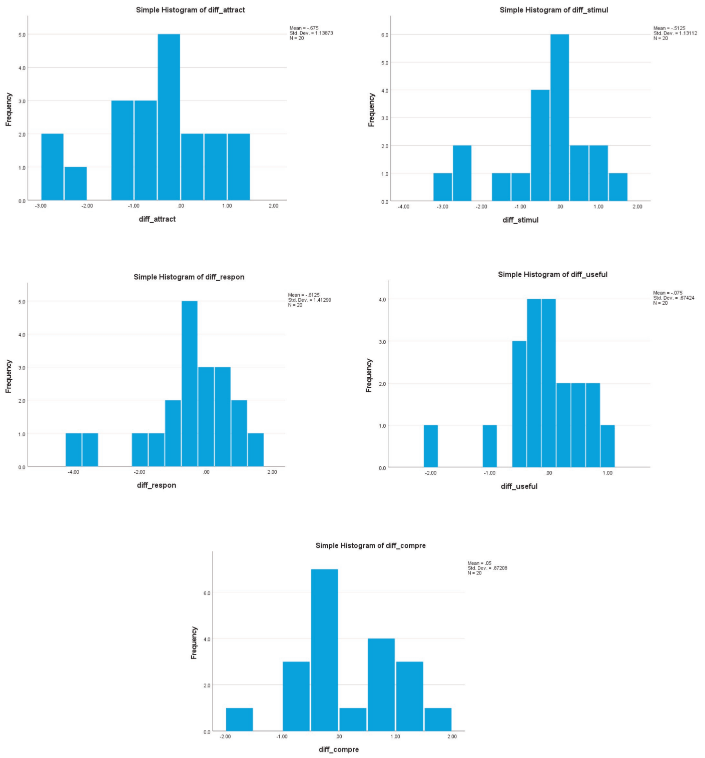
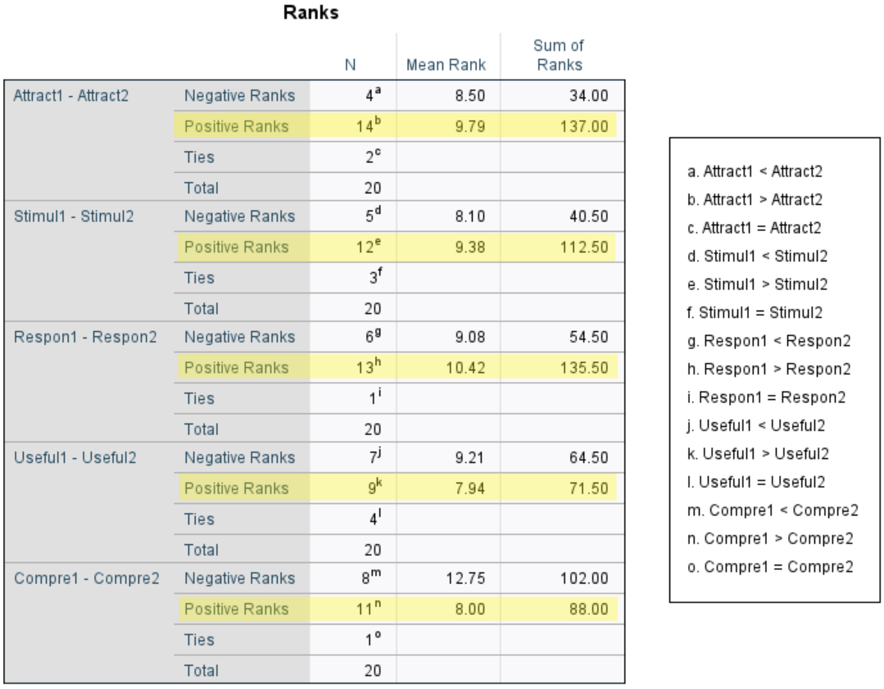


Figure 6.4: Distributional assumption for Wilcoxon test

The highlighted columns in Fig 6.5 show the number of participants that rated attractiveness of emotionally style voice (C1) higher than neutrally styled voice (C2). For the constructs of attractiveness, stimulation and response behaviour, more participants rated emotionally styled voice higher than a neutral voice. There is not a considerable difference for the rank test of usefulness and comprehensibility. Of the 20 participants recruited for this study, emotional styling was rated better by 14 participants for attractiveness, 12 for stimulation and 13 for response behaviour.



Ranks

		N	Mean Rank	Sum of Ranks
Attract1 - Attract2	Negative Ranks	4[a]	8.50	34.00
	Positive Ranks	14[b]	9.79	137.00
	Ties	2[c]		
	Total	20		
Stimul1 - Stimul2	Negative Ranks	5[d]	8.10	40.50
	Positive Ranks	12[e]	9.38	112.50
	Ties	3[f]		
	Total	20		
Respon1 - Respon2	Negative Ranks	6[g]	9.08	54.50
	Positive Ranks	13[h]	10.42	135.50
	Ties	1[i]		
	Total	20		
Useful1 - Useful2	Negative Ranks	7[j]	9.21	64.50
	Positive Ranks	9[k]	7.94	71.50
	Ties	4[l]		
	Total	20		
Compre1 - Compre2	Negative Ranks	8[m]	12.75	102.00
	Positive Ranks	11[n]	8.00	88.00
	Ties	1[o]		
	Total	20		

a. Attract1 < Attract2
b. Attract1 > Attract2
c. Attract1 = Attract2
d. Stimul1 < Stimul2
e. Stimul1 > Stimul2
f. Stimul1 = Stimul2
g. Respon1 < Respon2
h. Respon1 > Respon2
i. Respon1 = Respon2
j. Useful1 < Useful2
k. Useful1 > Useful2
l. Useful1 = Useful2
m. Compre1 < Compre2
n. Compre1 > Compre2
o. Compre1 = Compre2

Figure 6.5: Ranks for Wilcoxon test

The significance values are shown in Fig 6.1. It can be seen that there was statistical significance for the attractiveness (p <.05) construct. Thus, we can say that attractiveness of a emotionally styled voice interface had a statistically significant improvement when compared to a neutral voice interface in this study. The values for stimulation (p = .087 and response behaviour (p = .101) are not statistically significant (p >.05).

Test Statistics (a)

	Attract2 - Attract1	Stimul2 - Stimul1	Respon2 - Respon1	Useful2 - Useful1	Compre2 - Compre1
Z	-2.248b	-1.710b	-1.640b	-.183b	-.283c
Asymp. Sig. (2-tailed)	**0.025**	**0.087**	**0.101**	**0.855**	**0.777**

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

c. Based on negative ranks.

Table 6.1: Wilcoxon signed rank test statistics

6.1.2 Paired Sample T-Test

We conducted a paired sample T-test to measure the difference between the Attractiveness, Stimulation, Response Behaviour,Usefulness and Comprehensibility constructs, between condition 1 (Emotionally styled VA) and condition 2 (Neutral VA). Before conducting the T-test, we considered the four assumptions as recommended on Laerd Statistics[1].

1. The first assumption for the paired samples t-test to be valid is that dependent variable that is measured is to be continuous. The data was collected from the survey during the user testing in the form of a Likert scale (7-point scale). Even though Likert type data could be considered as ordinal data, the survey used in this study uses 4 Likert type items to produce a composite Likert. Such Likert scale data could be classified as interval scale items according to [5]. [5] further goes on to recommend that appropriate data analysis procedures for interval scale items include the Pearson's r, T-test, ANOVA, and regression procedures. Therefore, we used this as a basis to satisfy the first assumption.

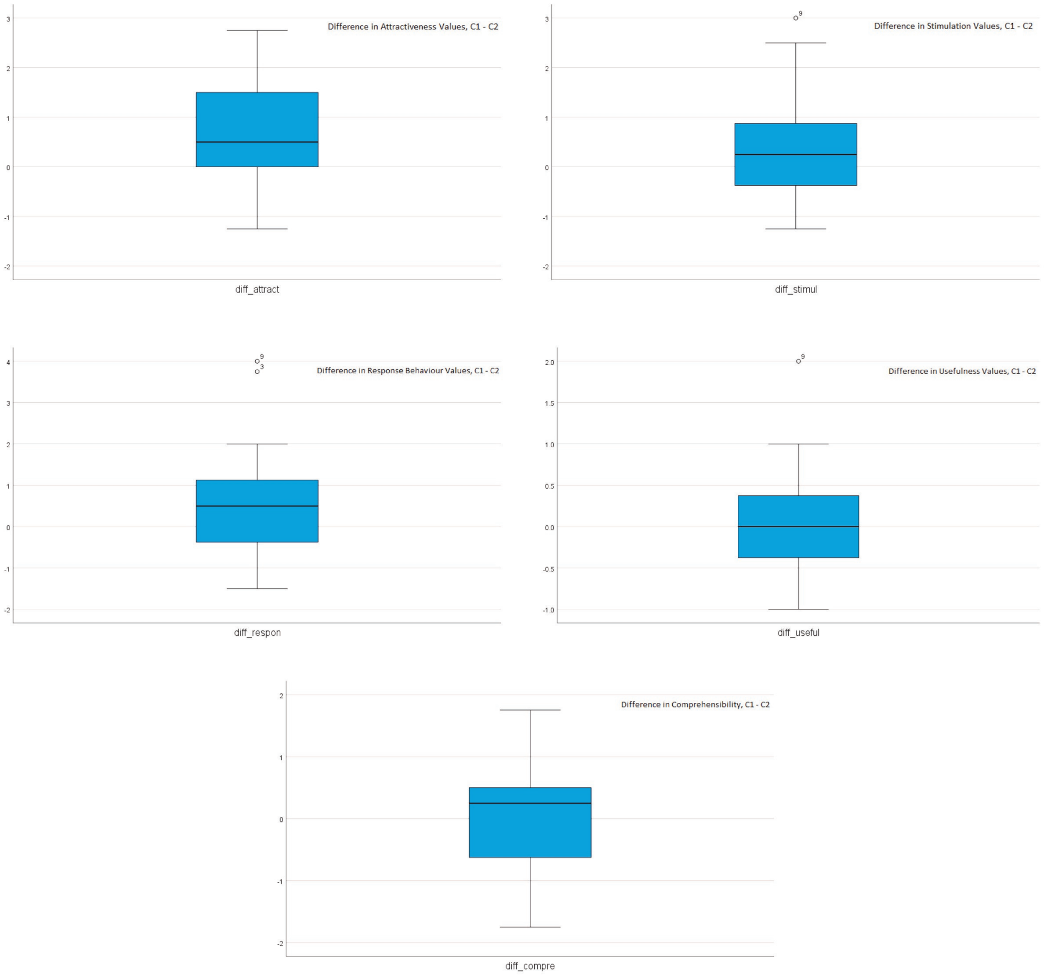


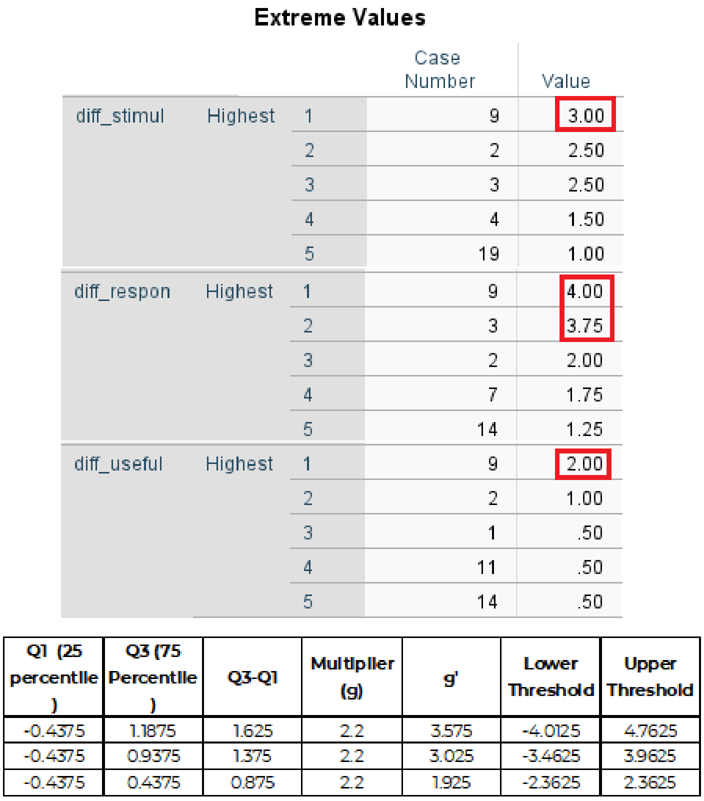
Figure 6.6: Boxplots to check outliers

2. The second assumption that needs to be considered for paired samples T-test is that the independent variable is dichotomous. In this study, the first category is the emotionally

[1] https://statistics.laerd.com/stata-tutorials/paired-t-test-using-stata.php

styled voice in condition 1 and the second category is the control or condition 2. Since all the participants experienced both these conditions, the second assumption for the test is also valid.

3. Assumptions number three states that there should not be significant outliers in the differences between the values of the two related groups. We plotted the difference of values between condition 1 and condition 2 for the five constructs using SPSS, as shown in Fig 6.6. As we take C1 minus C2, any differences observed here reflect an improvement of user experience in that particular construct. Two outliers were detected in the response behaviour construct and one outlier each, in stimulation and usefulness constructs. We investigated these 4 outliers and observed that they lie outside the inter-quartile range(IQR) rule when a multiplier equal to 1.5 is used. However, in Fig [19] it was suggested that a multiplier of 2.2 would be more valid. The values of these for outliers lie within the IQR of 2.2 as shown in 6.7. The upper and lower threshold limits along with the extreme values for case number 3 and 9 for responsive behaviour, case 9 for both stimulation and usefulness are displayed in Fig 6.7.



Extreme Values

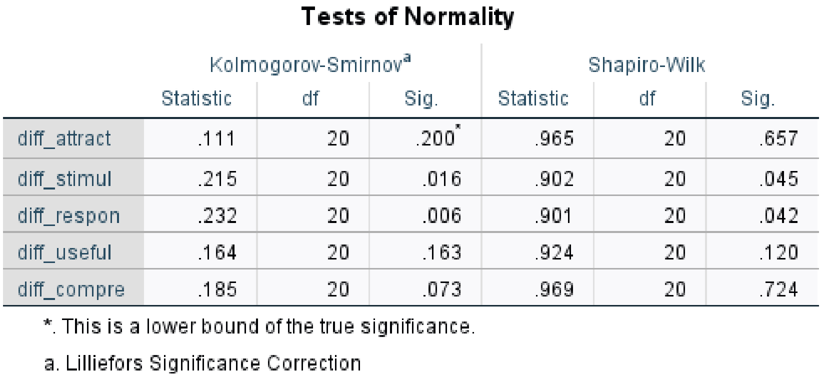
			Case Number	Value
diff_stimul	Highest	1	9	3.00
		2	2	2.50
		3	3	2.50
		4	4	1.50
		5	19	1.00
diff_respon	Highest	1	9	4.00
		2	3	3.75
		3	2	2.00
		4	7	1.75
		5	14	1.25
diff_useful	Highest	1	9	2.00
		2	2	1.00
		3	1	.50
		4	11	.50
		5	14	.50

Q1 (25 percentile)	Q3 (75 Percentile)	Q3-Q1	Multiplier (g)	g'	Lower Threshold	Upper Threshold
-0.4375	1.1875	1.625	2.2	3.575	-4.0125	4.7625
-0.4375	0.9375	1.375	2.2	3.025	-3.4625	3.9625
-0.4375	0.4375	0.875	2.2	1.925	-2.3625	2.3625

Figure 6.7: Outliers test using a new assumption for IQR multiplier

4. The fourth assumption is that the distribution of the difference. Using the difference of the values from the five constructs, we ran the Shapiro-Wilk test of Normality along with the Normal Q-Q plots. From the results of the Shaprio-Wilk as shown in Fig 6.8, it can be seen that for attractiveness (p = .657), usefulness (p = .120) and comprehensibility (p = .724) are normally distributed. Stimulation (p = .045) and response behaviour (p = .042) are not normally distributed (i.e. p <.05). In such cases, the Wilcoxon signed rank test, which was already conducted in6.1 is recommended when the normality assumption is not met. Nevertheless, according to [35],the paired sample T-test is robust to violations of normally

distributed data. Therefore, even though normality was violated for two of the constructs, we went ahead and conducted the T-test.



Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
diff_attract	.111	20	.200*	.965	20	.657
diff_stimul	.215	20	.016	.902	20	.045
diff_respon	.232	20	.006	.901	20	.042
diff_useful	.164	20	.163	.924	20	.120
diff_compre	.185	20	.073	.969	20	.724

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figure 6.8: Shapiro-Wilk normality test

Data shown here are mean ± standard deviation, unless otherwise stated. Participants scored emotionally styled (C1) voice interface higher as opposed to neutrally styled (C2) voice interface for the modules of attractiveness (C1 = 5.325 ± 1.267, C2 = 4.650 ± 0.897), stimulation (C1 = 4.963 ± 1.304, C2 = 4.525 ± 1.121), response behaviour(C1 = 5.025 ± 1.243, C2 = 4.413 ± 1.249), usefulness (C1 = 5.4125 ± 1.033, C2 = 5.338 ± 1.080). However, comprehensibility was scored lower (C1 = 4.825 ± 0.847, C2 = 4.875 ± 0.776).

The emotionally styled VA elicited in a higher rating by 0.675 (95% Confidence Interval (CI), 0.142 to 1.208) for attractiveness, 0.438 (95% CI, -0.102 to 0.977) for stimulation, 0.61250 (95% CI, -0.488 to 1.274) for response behaviour and 0.075 (95% CI, -0.241 to 0.391) for usefulness in this study. On the other hand, emotionally styled VA scored lower in the comprehensibility construct by -0.050 (95% CI, -0.458 to 0.358). The effect size was calculated dividing the mean difference between the two related groups with the standard deviation of this difference. According to [12], there is a large effect for attractiveness, stimulation, response behaviour and comprehensibility. However, usefulness only had a medium effect on the result of this study.

Similar results to the Wilcoxon Signed Rank Test were seen in the statistical significance of these five constructs as shown in Fig 6.2. Only attractiveness is statistically significant with (p <0.05). The values for stimulation and response behaviour are p=.106 and p=.068 which are both statistically insignificant (p >0.05) which is similar to what was seen in the Wilcoxon signed rank test in section 6.1.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Attractiveness1 - Attractiveness2	0.67500	1.13873	0.25463	0.14206	1.20794	2.651	19	0.016
Pair 2	Stimulation1 - Stimulation2	0.43750	1.15244	0.25769	-0.10186	0.97686	1.698	19	0.106
Pair 3	ResponseBehaviour1 - ResponseBehaviour2	0.61250	1.41299	0.31595	-0.04880	1.27380	1.939	19	0.068
Pair 4	Usefullness1 - Usefullness2	0.07500	0.67424	0.15077	-0.24056	0.39056	0.497	19	0.625
Pair 5	Comprehensibility1 - Comprehensibility2	-0.05000	0.87208	0.19500	-0.45815	0.35815	-0.256	19	0.800

Table 6.2: Paired sample T test significance values

From both the Wilcoxon signed rank test and the Paired sample T-test it can be seen that there is a statistically significant difference between the attractiveness constructs for EVA when compared to CVA. Therefore it can be said that participants rated an emotionally styled voice higher in terms of attractiveness. However, the other four constructs of user experience did not show statistically significance.

6.2 Qualitative Data

A total of 21 participants were interviewed after completing both conditions of the experiment. The responses from the interview were audio recorded. We asked a total of 10 questions as shown A.2.1, with a few follow up questions based on participants responses. On an average, each interview lasted for about 15 minutes.

Procedure

All the transcribed data was collected into a word document in the form of text. We split this data into four parts and subjected it to thematic analysis.

We based our qualitative analysis on grounded theory from [17]. We chose grounded theory as the qualitative data analysis method because it allowed us to start theming the parts of the interviews (P1 to P5) while still conducting the user study with more participants. Another reason for choosing grounded theory is that it adopts a neutral view of without the need to make assumptions about the data. Finally, grounded theory enables us to link data to existing theory, while also simultaneously using the qualitative and quantitative data together[11].

The analysis started with summarizing and coding participant responses in blocks of five participants each. The summarized text consists of direct quotes from participants, implied meaning from their replies and their preferences. We used the constant comparison method [55] to compare each themed block with data from other themed blocks.

Results

Out of 20 participants, 3 participants did not notice any changes between the two conditions. Another 3 noticed subtle changes in the details of the content between the two conditions. The participant who experienced internet connectivity issues during the test was omitted from this list, since a considerable delay was noticed by them in the first condition when compared to the second condition (due to slow internet speed). 6 out of 20 participants stated that the sad or negative emotion of EVA felt artificial. On hearing EVA and CVA back-to-back, 18 out of 21 participants preferred EVA.

When asked to choose between a positively versus a negatively styled emotion, 18 out of 21 participants chose positively styled emotions. 11 out of 21 users wanted the VA to deliver negative content in a neutral voice. 5 wanted it in a negative style and the remaining 5 did not like the implementation of the negative emotion in the current prototype.

Proactivity received mostly positive feedback, albeit with a caveat. Participants had reservations about proactivity as they did not want to be interrupted in complex driving situations. Contextual proactivity was highly rated. Of the 21 total participants 16 of them stated that even with proactivity, the final decision should be left to them. They did not want the system to make decisions without their consent (as seen in the parking situation). Non-sports fans suggested that instead of sports, they would be open towards other contextual proactivity in topics of their interest. Proactive parking was the highest rated, with 17 out of 21 participants picking it. Users expressed that their willingness to accept a proactive VA would depend on factors such as their mood, time of day and context. Two participants did not like the idea of affective systems. Both of them were against VAs showing emotions. They expressed that future systems with personalities, could cause unnatural attachment of humans to Artificial Intelligence (AI) systems.

Chapter 7

Discussion

In this section we discuss our findings from the data collected during the user study. We summarize and interpret the results obtained from the previous chapter 6. We also discuss our findings, from the qualitative interviews.

The data collected from the questionnaire shows the means for user experience; i.e. 4 out of the 5 constructs (attractiveness, stimulation, response behaviour and usefulness) were higher for emotionally styled voice system when compared to neutral styled voice. These findings show potential agreement with our primary hypothesis that an emotionally styled voice system would have better user experience. However, just comparing the means alone would be insufficient to accept our hypothesis. The data analysis using both the Wilcoxon and paired sample T-test in chapter 6 show a statistically significant mean increase in the scores for the user experience construct of attractiveness for the emotionally styled voice assistant (EVA) over the neutrally styled (or) control voice assistant (CVA), t(19) = 2.651, p <.05, d = 1.138. However, the other constructs of the UEQ+ questionnaire used in this study, did not elicit a statistically significance between the EVA and the CVA conditions.

7.1 Emotional Styling Of Voice

Participants were asked if they noticed any differences between the two voice assistants. Out of 20 participants, 14 participants noticed a difference between the two voice styles. Of the 6 participants who did not identify a difference in voice style, we found that 5 are not regular users of voice assistants (4 use it once a month, 1 uses VAs less than once a month). Thus we could postulate that novice users of voice interfaces would be more likely to miss the subtle nuances in voice. These participants could have just been overwhelmed by the novelty effect [33], and therefore failed to notice the difference in voice styling. However, even though they did not notice a difference in emotions within the voice, a few of them used descriptive words such as “conversational”, “natural” and “coherent” when referring to EVA.

Of the 14 participants that noticed a difference in the voice style between the two VAs in the experiment, 10 participants described EVA with the words “friendly”, “informal”, “conversational”, “likable”, “smoother”, “approachable”, “energised” and “lively”. Therefore, the participants who noticed variation in voice during the user study, a majority of them used positive terms to describe their interaction with EVA. The general consensus was that EVA felt like it had a personality and participants could “sense” it’s happiness during the positive news. However, out of these 14 participants, 6 of them stated that the sad or negative emotion of EVA felt “artificial” or “off”. P04 stated that “EVA felt artificial and very sad. The fact that it started off with telling me about sad news (weather prompt) at the beginning of the ride , put me off for the rest of the interaction". Conversely, these 14 participants described CVA as "mechanical", "robotic", "informative", "scripted", "plain", "dull", "monotonous" and "inorganic". The word pairs informative for CVA and conversational for EVA suggest that introducing appropriate emotional styling could improve voice interaction between humans and voice systems. It also suggests that users would be more likely to assume an emotionally styled system as a human co-driver when compared to a neutral system.

The researcher then played the sports voice prompt with positive content, with positively styled, followed by the same prompt with the neutrally styled voice for participants to hear back-to-back. This was done to work as a manipulation check to see if participants subjective evaluation of the displayed emotion, matched the intended emotion of the voice style. When participants experienced the back-to-back playback, 18 out of 21 preferred EVA with the lively voice style over a neutrally styled CVA. Of the 3 participants who preferred CVA over EVA, 2 did not notice a difference and 1 (P04) of them was discouraged by the artificial negative voice of EVA. EVA was preferred by most participants as they felt it was more human-like, enthusiastic and excited about the fact that their favourite sports team won the game. Participants also stated that it felt like EVA cares about them. Therefore, users were more willing to answer to EVA when compared to CVA.

Participants were then asked about their thoughts on VAs displaying emotions. A majority of participants expressed that having emotions, made the VA feel more like a co-driver, rather than an assistant. They would be more likely to build a connection and engage more with a VA that is capable of showing appropriate emotions. However, many of them expressed their concern over how the assistant would appropriately match the message content to emotional style. P13 said "I felt that the lockdown emotion was overplayed so I would get annoyed and call the system stupid". P10 experienced a incorrect emotional response during their interaction. When P10 said "I would spend the two weeks (during lockdown) alone at home and not do anything", an incorrect reply prompt was triggered saying "That sounds like a good plan" in a lively voice. This inappropriate matching of emotions by the EVA caused frustration and annoyed the user more than having a neutral voice for this incorrect message content. Therefore it is vital that the emotional style matching of the prompt is in accordance not only with the content of the message, but also with the user tastes or preferences. For example, if a user wants another lockdown so that he can spend time at home with his family, a VA which uses a negative tone to deliver this message would annoy the user as its emotions do not match the users expectations. A few participants expressed the desire to know more about the VAs personality and likes. A few other users suggested that intensity of emotions should vary with content in accordance with their preferences. Participants also wanted to ask funny questions to see if they can confuse, find limitations in the systems capabilities to make fun of or bully the system [40].

Two users responded with an interesting observation about the ethically viability of affective systems. They said that as the first step, VAs could show emotions, later they might develop a personality, after which a relationship between the user and the system is formed. They further went on to say that such a relationship might make humans too attached and dependant on such affective systems. This could mean that when users are unable to interact with their affective VA "friend" anymore (due to an update or if they use a device without affective capabilities) they could experience similar feeling to that of losing a friend or a partner. This raises psychological and ethical questions about possible relationships with affective agents [25] and could potentially have harmful long term effects.

Within the emotionally styled VA, the positive were preferred over negative emotions by 18 out of the 21 participants. A few reasoned that positive styled voice would give them "good vibes", keeping them in a good mood (13/21). Other reasoned that a prompt with negative style itself would already contain bad information, they wouldn't want their mood to be ruined further by the sadness in the voice (11/21). They were asked whether they'd prefer voice styling with only positively matched content, only negatively matched content or both positive and negatively matched content or entirely neutral styled. 11 out of 21 said that they would prefer having bad news delivered to them in a neutral style, so that it feels like ripping off a band-aid and not sugarcoating bad news. 5 out of 21 said they would prefer negative styled voice for bad news. The remaining 5 stated that they felt the implementation of negative voice in this study felt artificial, and would not be prefer it at all. A few participants did express that just because the VA uses a emotional voice style does not mean that it is expressing emotions. However, P06 said "A human designed this emotional style, so I know that it is not being faked by a VA".

When asked about their willingness to use an emotional VA, 17 participants expressed that they would, as the interaction would become more human-like. P10 was concerned about appropriate matching of emotions to user expected emotions. P08 and P09 wanted the system to not just be emotional, but also have its own personality. Participants state that correct matching of emotion to the message content would make them want to interact for longer periods of time with an emotional voice assistant.

7.2 Proactivity of Voice Assistant

A majority of the participants stated that the VA should be proactive only when they are not paying attention to the road. They expressed safety concerns about proactive notifications by the VA during complicated driving situations, which is in agreement with the research conducted in [46]. 4 users found the notification beep irritating and loud in the starting. However, they stated that they got comfortable to the beep with time. Most users expressed their liking towards contextual proactivity and found it very interesting. A couple of users however stated that the sports teams score might come off as a spoiler, therefore the VA should ask the user, before telling them the score. Participants also said that they would want to be in control of the content and time the system is proactive. An important observation was that almost all the users stated their frustration with the system making decisions for them, especially in the parking scenario(we had to remove user choice to keep the flow the same between the two conditions, also explained in detail in 5.3). According to them, the system could be proactive, but should always leave the decision making to the user. Finally, non sports fans did not find the sports use-case helpful and expressed that they liked the contextual nature of it, but would prefer it to be proactive with topics of their interests in the surrounding. P06 said "I was noticing that there was a man running with yellow shoes, and that really caught my attention. So maybe, sensing my gaze towards the shoes, the assistant tells me more about the product".

The proactive parking use-case was preferred by 17 participants. The reason given by participants for choosing parking is because it is usually the main goal when driving to a destination. However, they state that they should be given the choice to decide the parking location. Of the 4 participants who did not have parking as their first choice, 2 expressed that they do not drive often, so they do not need to think about parking often. Non-sports fans expressed that proactivity should be personalised to their interests. Weather was also rated well by most participants, but a few expressed the desire to ask the VA follow up questions to find more details about the forecast. Lockdown news had mixed reviews, with a few users stating that they preferred their phone to give them such news rather than a VA in an automobile.

Users expressed their willingness to have a proactive VA only if were contextually correct i.e. both content and time of proactivity. Further on, users were more willing to use an intelligent system which is capable of showing emotions, when compared to a proactive VA. The willingness to use proactivity also depends on the mood of the user and almost all the participants state that system needs to be adaptable to their preferences.

7.3 Data Privacy

Another important factor that needs to be addressed when dealing with personalisation of intelligent voice systems is privacy. Personalisation a VA could require data collection, in order to provide recommendations and customized content tailored for individual users. The collection, usage, processing and storage of data brings with it privacy concerns about user data. Users must be aware and given a choice as to how their data would be processed and stored by the system. In the current project we processed the data in accordance to GDPR regulations [53]. For a VA that is used daily, a lot of data would be generated and stored by a system. A proactive system would also need to predict the right time, therefore it must always be "listening". Thus, when it comes to proactivity of systems, privacy would be a key challenge to overcome. In case of shared cars or car pooling, the system might need to be aware of other users in the car should not share sensitive information about the primary user. It should also be aware of the contextual situation and determine the occupants in the car to decide which information to convey. When explicitly asked, users did not have a positive attitude towards such proactive system's constant data collection policy. Therefore data privacy should to be addressed by potentially giving the user control over how the system collects, uses, processes and stores their data.

7.4 Limitations of this study

As we conducted a remote user study, there are a few limitations to our method. The limitations are as follows:

1. Even though we used a video to provide the illusion of users riding in an autonomous car, this is not equivalent to a realistic driving situation. Remote testing could have influenced user ratings differently when compared to a realistic driving environment.
2. The VA in the study was deployed using a Wizard-of-OZ (WoZ) setup. We could not implement all the functionalities that are seen as commonplace with VAs, such as voice search to provide answers to user queries, changing in-car functions, sending text, making calls etc. By informing users that they cannot ask the system any questions, the conversation does not remain natural (as with a co-driver).
3. Leaving the users to interact with the system unmoderated could have elicited more natural responses from them. However, this was not possible in the current study, as developing a system that is capable of handling complicated situations would take considerable time and effort.
4. Using the usefulness module of the UEQ+ gave us almost the same rating for both conditions by users. The usefulness construct measures the system's ability to make it easier for users to reach goals, save time or improve productivity. Therefore, this construct measures how important the scenarios are in the user study. Since both the conditions in our study had the same scenarios, there was negligible difference in the usefulness ratings by participants. In hindsight, we should have measured the usefulness only once, after both the conditions.
5. All the users experienced the sports scenario, irrespective of their preferences. We could not collect individual preference before the study, to customize the scenarios to specific users. Personalising the scenario based on user preferences would mean that different users would experience different situations. Not only would this make the experiment setup complicated, but also make it difficult to quantify and compare between similar situations for different users.

Chapter 8

Conclusion

The goal of this project was to investigate emotionally intelligent voice interaction systems. We started with a broad topic of personalisation of a proactive voice assistant. User research was done in order to gain insights into usage patterns and scenarios of such systems. We conducted multiple focus groups and interviews to collect user data and develop scenarios for our user study. We collected multiple use-cases and prioritized them into the most valuable scenarios in relation to this project. After identifying key use cases, we developed a persona to act as the basis for the conversation design of the study. A experiment scenario was then created in accordance with the research goal. Iterative changes were made to this scenario by keeping it in within multiple factors such as relevance to the research question, duration of the study, technical possibilities etc. We narrowed down the scope of this project. Our primary research goal was to investigate the effect of emotional voice styling on user experience of proactive voice systems.

After developing the scenario, we designed an experiment procedure. We carried out four pre-studies to identify issues with the study. These issues were resolved and the study was posted on various platforms within the Eindhoven University of Technology to start recruiting participants. We conducted an experiment with a total of 21 participants in a remote user study. Qualitative results shown that the means of the emotionally styled voice were higher for attractiveness, stimulation and response behaviour when compared to a neutrally styled voice. Statistical tests on the quantitative data confirmed that the values of attractiveness between the two conditions was statistically significant ($p < .05$). These findings reject the null hypothesis and supports our alternate hypothesis which states that users would find an emotionally styled voice assistant to have better user experience over a neutrally styled assistant. Users rated the usefulness of the proactive system in the current project highly (mean = 1.54, on the scale -3 to +3). Users also found the emotional voice assistant to behave like a human co-driver, when compared to the neutral system.

Emotional styling could improve the user experience, acting as a first step towards having a natural conversation with voice assistants. However, it is essential that emotional style should be appropriately matched to the content of the message, based user preferences. Contextual proactivity is desired by users, but with a few reservations. Proactive systems should be adaptive in nature, starting off with only essential interruptions and then adapting to the user's choice.

8.1 Future Work

An extension to the current study would be to explore the effects of positive versus negative emotional styling. Furthermore, the magnitude of each emotion could be varied according to user preferences. However, such systems would need considerable amount of user data to enable such personalisation. As seen from the results in section 6, participants expressed that emotional styling would more appropriate in situations of personal interest. Therefore, another approach would be to give explicit choice to users, to pick topics for which they want a VA to be emotionally styled. Such systems would not only be personalised content wise, but also from the perspective of emotional expression that conforms to the user's subjective evaluation of the situation.

Further studies could be conducted with a VA that is both proactive and reactive in contextually personalised situations. Such research could explore the domain of proactivity with a more adaptable VA. Such a system could implement both current capabilities of VAs (ability to respond to queries) along with proactivity to investigate user perceptions.

Personal Reflection

The journey taken during my master thesis was challenging, yet intellectually satisfying at the same time. I learnt many things along the way and found myself in uncharted waters a few times. Thankfully, my supervisors, both internal and external, were very patient and helpful in guiding this project to fruition.

The main takeaway from this thesis for me was understanding the user centered research process. I believe that my ability to recognise the subtle nuances in user feedback have improved. Interacting with participants with different outlooks towards interactive technology within cars, gave me valuable insights about user behaviour. I also learnt that since this is a cyclical process, the more perspectives we get for an idea, the greater it's impact on the final outcome. Finally, this thesis has taught me that a focused mindset towards the end goal can overcome difficulties in even in the most uncertain of times.

Bibliography

[1] Abras, C., Maloney-Krichmar, D., Preece, J., et al. User-centered design. *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications* (2004). 10, 14

[2] Andrist, S., Mutlu, B., and Tapus, A. *Look Like Me: Matching Robot Personality via Gaze to Increase Motivation.* Association for Computing Machinery, New York, NY, USA, 2015, p. 36033612. 7

[3] Bernhaupt, R., Murko, C., Pottier, G., and Battut, A. User acceptance of emotion-aware mood-improving voice assistants. *International Broadcasting Convention* (October 2019). 7

[4] Bickmore, T. W., and Picard, R. W. Establishing and maintaining long-term human-computer relationships. *ACM Trans. Comput.-Hum. Interact.* (2005). 7

[5] Boone, H. N., and Boone, D. A. Analyzing likert data. *Journal of extension 50*, 2 (2012), 1–5. 46

[6] Braun, M., and Alt, F. *Identifying Personality Dimensions for Characters of Digital Agents.* Springer International Publishing, Cham, 2020, pp. 123–137. 7, 8, 9

[7] Braun, M., Li, J., Weber, F., Pfleging, B., Butz, A., and Alt, F. What if your car would care? exploring use cases for affective automotive user interfaces. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services* (New York, NY, USA, 2020), MobileHCI '20, Association for Computing Machinery. 7, 14

[8] Braun, M., Mainz, A., Chadowitz, R., Pfleging, B., and Alt, F. At your service: Designing voice assistant personalities to improve automotive user interfaces. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2019), CHI 19, Association for Computing Machinery. 6, 7, 8, 9, 25, 31

[9] Breazeal, C., and Aryananda, L. Recognition of affective communicative intent in robot-directed speech. *Autonomous robots 12*, 1 (2002), 83–104. 6

[10] Brown, P. F. The acoustic-modeling problem in automatic speech recognition. Tech. rep., CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF COMPUTER SCIENCE, 1987. 4

[11] Charmaz, K., and Belgrave, L. L. Grounded theory. *The Blackwell encyclopedia of sociology* (2007). 49

[12] Cohen, J. *Statistical power analysis for the behavioral sciences.* Academic press, 2013. 48

[13] Dimock, M. Defining generations: Where millennials end and generation z begins. *Pew Research Center* (2019). 32

[14] Ekman, P. An argument for basic emotions. *Cognition & emotion 6*, 3-4 (1992), 169–200. 8

[15] Etikan, I., Musa, S. A., and Alkassim, R. S. Comparison of convenience sampling and purposive sampling. *American journal of theoretical and applied statistics 5*, 1 (2016), 1–4. 34

[16] Fugard, A. J., and Potts, H. W. Supporting thinking on sample sizes for thematic analyses: a quantitative tool. *International Journal of Social Research Methodology 18*, 6 (2015), 669–684. 27

[17] Glaser, B. G., and Strauss, A. L. *Discovery of grounded theory: Strategies for qualitative research.* Routledge, 2017. 49

[18] Göker, A., and Myrhaug, H. I. User context and personalisation. In *ECCBR Workshops* (2002), vol. 2002, pp. 1–7. 7

[19] Hoaglin, D. C., and Iglewicz, B. Fine-Tuning Some Resistant Rules for Outlier Labeling. *Journal of the American Statistical Association 82*, 400 (1987), 1147–1149. 47

[20] Klein, A. M., Hinderks, A., Schrepp, M., and Thomaschewski, J. Construction of ueq+ scales for voice quality: Measuring user experience quality of voice interaction. In *Proceedings of the Conference on Mensch Und Computer* (New York, NY, USA, 2020), MuC '20, Association for Computing Machinery, p. 15. 32

[21] Klein, A. M., Hinderks, A., Schrepp, M., and Thomaschewski, J. *Construction of UEQ+ Scales for Voice Quality: Measuring User Experience Quality of Voice Interaction.* Association for Computing Machinery, New York, NY, USA, 2020, p. 15. 32

[22] Kun, A., Paek, T., and Medenica, Z. The effect of speech interface accuracy on driving performance. In *Eighth Annual Conference of the International Speech Communication Association* (01 2007), vol. 4, pp. 1326–1329. 5

[23] Large, D. R., Burnett, G., Anyasodo, B., and Skrypchuk, L. Assessing cognitive demand during natural language interactions with a digital driving assistant. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (New York, NY, USA, 2016), AutomotiveUI 16, Association for Computing Machinery, p. 6774. 5

[24] Laugwitz, B., Held, T., and Schrepp, M. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and usability engineering group* (2008), Springer, pp. 63–76.

[25] Levy, D. *Love and sex with robots: The evolution of human-robot relationships.* New York, 2009. 52

[26] MacLean, A., Young, R. M., Bellotti, V. M. E., and Moran, T. P. Questions, options, and criteria: Elements of design space analysis. *Hum.-Comput. Interact.* (1991).

[27] Meuleman, B., and Scherer, K. R. Nonlinear appraisal modeling: An application of machine learning to the study of emotion production. *IEEE Transactions on Affective Computing 4*, 4 (2013), 398–411. 7

[28] Miksik, O., Munasinghe, I., Asensio-Cubero, J., Bethi, S. R., Huang, S., Zylfo, S., Liu, X., Nica, T., Mitrocsak, A., Mezza, S., et al. Building proactive voice assistants: When and how (not) to interact. *Retrieved from arxiv preprint, arxiv:2005.01322* (2020).

[29] Nass, C., Jonsson, I.-M., Harris, H., Reaves, B., Endo, J., Brave, S., and Takayama, L. Improving automotive safety by pairing driver emotion and car voice emotion. In *CHI 05 Extended Abstracts on Human Factors in Computing Systems* (New York, NY, USA, 2005), CHI EA 05, Association for Computing Machinery, p. 19731976. 7, 9

[30] Nass, C., Jonsson, I.-M., Harris, H., Reaves, B., Endo, J., Brave, S., and Takayama, L. Improving automotive safety by pairing driver emotion and car voice emotion. In *CHI 05 Extended Abstracts on Human Factors in Computing Systems* (New York, NY, USA, 2005), CHI EA 05, Association for Computing Machinery, p. 19731976.

[31] Nass, C., Moon, Y., Fogg, B. J., Reeves, B., and Dryer, C. Can computer personalities be human personalities? In *Conference Companion on Human Factors in Computing Systems* (New York, NY, USA, 1995), CHI 95, Association for Computing Machinery, p. 228229. 7, 31

[32] Peissner, M., Doebler, V., and Metze, F. Can voice interaction help reducing the level of distraction and prevent accidents. *Meta-Study Driver Distraction Voice Interaction* (2011), 24. 5

[33] Poppenk, J., Köhler, S., and Moscovitch, M. Revisiting the novelty effect: When familiarity, not novelty, enhances memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition 36*, 5 (2010), 1321. 51

[34] Press Release, . Bmw group intelligent personal assistant 'hey bmw', 2018. Retrieved on June 12, 2020. 2

[35] Rasch, D., and Guiard, V. The robustness of parametric statistical methods. *Psychology Science 46* (2004), 175–208. 47

[36] Reagan, I., and Cicchino, J. Do Not Disturb While Driving Use of cellphone blockers among adult drivers. *Safety Science 128* (2020), 104753. 5

[37] ROGERS, E. M. Diffusion of innovations. *Farmers and researchers: How can collaborative advantages be created in participatory research and technology development?* (2007), 37.

[38] SAE International, . Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles, 2018. Retrieved on May 22,2020. 3

[39] Sahin, I. Detailed review of rogers' diffusion of innovations theory and educational technology-related studies based on rogers' theory. *Turkish Online Journal of Educational Technology-TOJET* (2006). 32

[40] Salvini, P., Ciaravella, G., Yu, W., Ferri, G., Manzi, A., Mazzolai, B., Laschi, C., Oh, S.-R., and Dario, P. How safe are service robots in urban environments? bullying a robot. In *19th International Symposium in Robot and Human Interactive Communication* (2010), IEEE, pp. 1–7. 52

[41] Scherer, K. R. Appraisal theory. In *Handbook of cognition and emotion.* John Wiley & Sons Ltd, New York, NY, US, 1999, pp. 637–663. 8

[42] Schmidt, M., and Braunger, P. A survey on different means of personalized dialog output for an adaptive personal assistant. In *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization* (New York, NY, USA, 2018), UMAP 18, Association for Computing Machinery, p. 7581. 6, 8

[43] Schmidt, M., Minker, W., and Werner, S. User acceptance of proactive voice assistant behavior. In *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2020* (2020), R. Bck, I. Siegert, and A. Wendemuth, Eds., TUDpress, Dresden, pp. 18–25. 6

[44] Schmidt, M., Stier, D., Werner, S., and Minker, W. Exploration and assessment of proactive use cases for an in-car voice assistant. In *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2019* (2019), P. Birkholz and S. Stone, Eds., TUDpress, Dresden, pp. 148–155. 6, 9, 14

[45] Schrepp, M., and Thomaschewski, J. Handbook for the modular extension of the user experience questionnaire. *All you need to know to apply the UEQ+ to create your own UX questionnaire. Google Scholar Google Scholar Cross Ref Cross Ref* (2019). 32

[46] Semmens, R., Martelaro, N., Kaveti, P., Stent, S., and Ju, W. Is now a good time? an empirical study of vehicle-driver communication timing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2019), CHI 19, Association for Computing Machinery. 6, 9, 53

[47] Shi, Y., Yan, X., Ma, X., Lou, Y., and Cao, N. *Designing Emotional Expressions of Conversational States for Voice Assistants: Modality and Engagement.* Association for Computing Machinery, New York, NY, USA, 2018, p. 16. 8

[48] Sodnik, J., Dicke, C., Tomažič, S., and Billinghurst, M. A user study of auditory versus visual interfaces for use while driving. *Int. J. Hum.-Comput. Stud. 66*, 5 (May 2008), 318332. 5

[49] Strauss, P., and Minker, W. *Proactive Spoken Dialogue Interaction in Multi-Party Environments.* Springer, 2010.

[50] Tapus, A., and Mataric, M. J. Socially assistive robots: The link between personality, empathy, physiological signals, and task performance. In *AAAI spring symposium: emotion, personality, and social behavior* (2008), pp. 133–140. 7

[51] Underwood, G., Chapman, P., Wright, S., and Crundall, D. Anger while driving. *Transportation Research Part F: Traffic Psychology and Behaviour* (1999). 8

[52] Voicebot.ai. In-car voice assitant consumer adoption report. Tech. rep., VoiceBot.ai, Jan 2020. Retrieved April 25, 2020. 2, 4

[53] Voigt, P., and Bussche, A. v. d. *The EU General Data Protection Regulation (GDPR): A Practical Guide*, 1st ed. Springer Publishing Company, Incorporated, 2017. 54

[54] Völkel, S. T., Schödel, R., Buschek, D., Stachl, C., Winterhalter, V., Bühner, M., and Hussmann, H. Developing a personality model for speech-based conversational agents using the psycholexical approach. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020). 8

[55] Walker, D., and Myrick, F. Grounded theory: An exploration of process and procedure. *Qualitative health research 16*, 4 (2006), 547–559. 49

[56] Watson, D., and Tellegen, A. Toward a consensual structure of mood. *Psychological bulletin 98*, 2 (1985), 219. 7

[57] Weber, D., Shirazi, A. S., and Henze, N. Towards smart notifications using research in the large. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (New York, NY, USA, 2015), MobileHCI 15, Association for Computing Machinery, p. 11171122. 6

[58] Zepf, S., Dittrich, M., Hernandez, J., and Schmitt, A. Towards empathetic car interfaces: Emotional triggers while driving. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2019), CHI EA '19, Association for Computing Machinery. 8

Appendix A

Appendix

A.1 User Research

A.1.1 Focus group with Experts

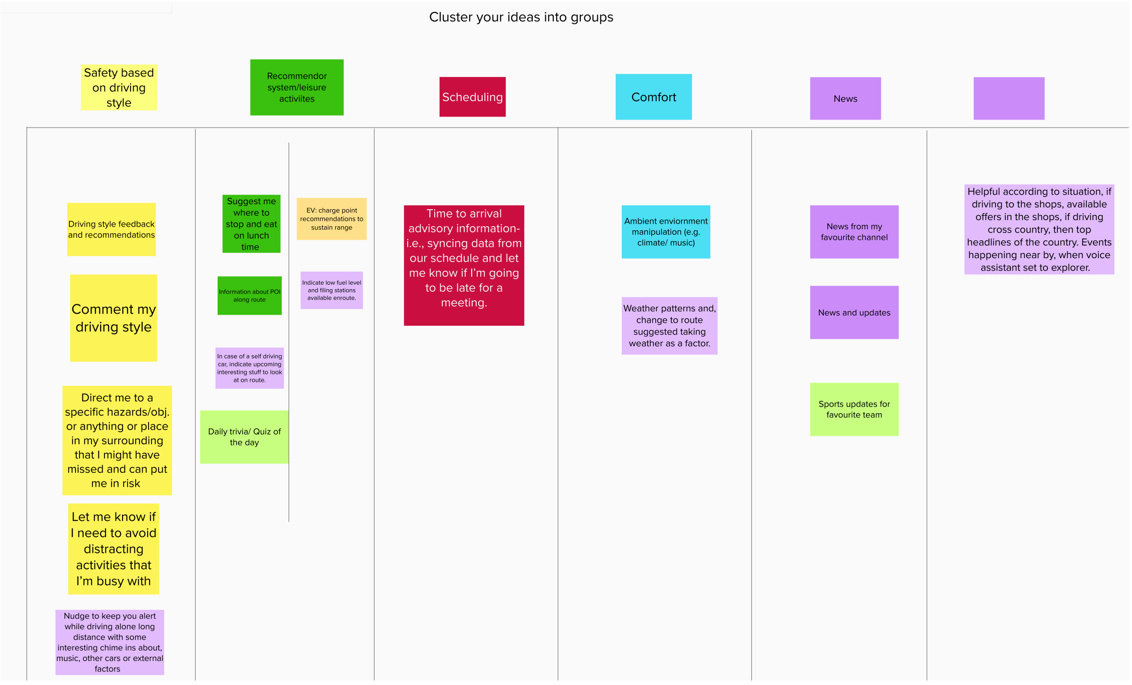
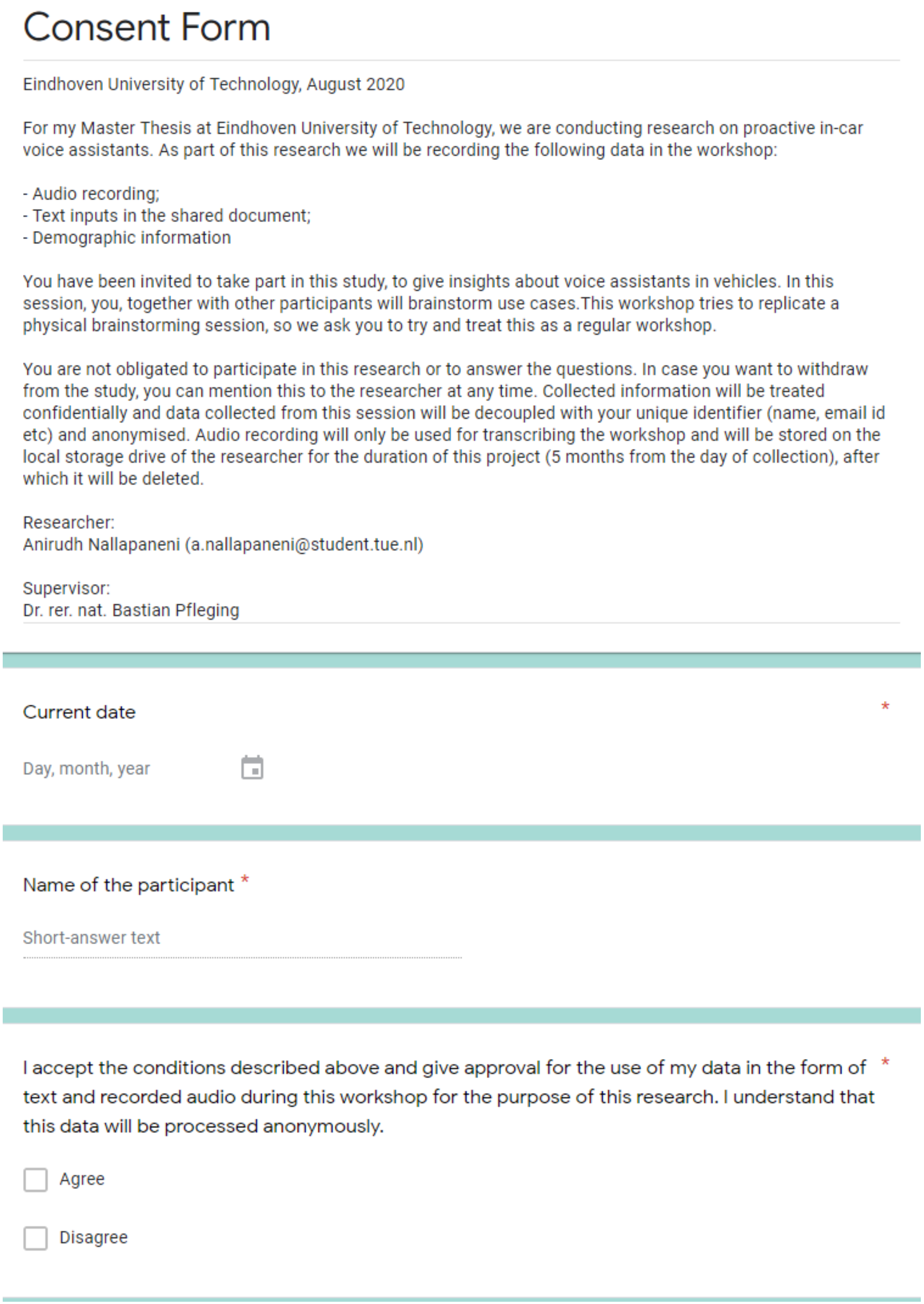

Cluster your ideas into groups
Safety based on driving style
Recommendor system/leisure activiites
Scheduling
Comfort
News
Driving style feedback and recommendations
Comment my driving style
Direct me to a specific hazards/obj. or anything or place in my surrounding that I might have missed and can put me in risk
Let me know if I need to avoid distracting activities that I'm busy with
Nudge to keep you alert while driving alone long distance with some interesting chime ins about, music, other cars or external factors
Suggest me where to stop and eat on lunch time
Information about POI along route
In case of a self driving car, indicate upcoming interesting stuff to look at on route.
Daily trivia/ Quiz of the day
EV: charge point recommendations to sustain range
Indicate low fuel level and filing stations available enroute.
Time to arrival advisory information- i.e., syncing data from our schedule and let me know if I'm going to be late for a meeting.
Ambient enviornment manipulation (e.g. climate/ music)
Weather patterns and, change to route suggested taking weather as a factor.
News from my favourite channel
News and updates
Sports updates for favourite team
Helpful according to situation, if driving to the shops, available offers in the shops, if driving cross country, then top headlines of the country. Events happening near by, when voice assistant set to explorer.

Figure A.1: Clustered ideas from Focus Group with Experts

A.1.2 Consent form for focus groups



Consent Form

Eindhoven University of Technology, August 2020

For my Master Thesis at Eindhoven University of Technology, we are conducting research on proactive in-car voice assistants. As part of this research we will be recording the following data in the workshop:

- Audio recording;
- Text inputs in the shared document;
- Demographic information

You have been invited to take part in this study, to give insights about voice assistants in vehicles. In this session, you, together with other participants will brainstorm use cases.This workshop tries to replicate a physical brainstorming session, so we ask you to try and treat this as a regular workshop.

You are not obligated to participate in this research or to answer the questions. In case you want to withdraw from the study, you can mention this to the researcher at any time. Collected information will be treated confidentially and data collected from this session will be decoupled with your unique identifier (name, email id etc) and anonymised. Audio recording will only be used for transcribing the workshop and will be stored on the local storage drive of the researcher for the duration of this project (5 months from the day of collection), after which it will be deleted.

Researcher:
Anirudh Nallapaneni (a.nallapaneni@student.tue.nl)

Supervisor:
Dr. rer. nat. Bastian Pfleging

Current date *

Day, month, year

Name of the participant *

Short-answer text

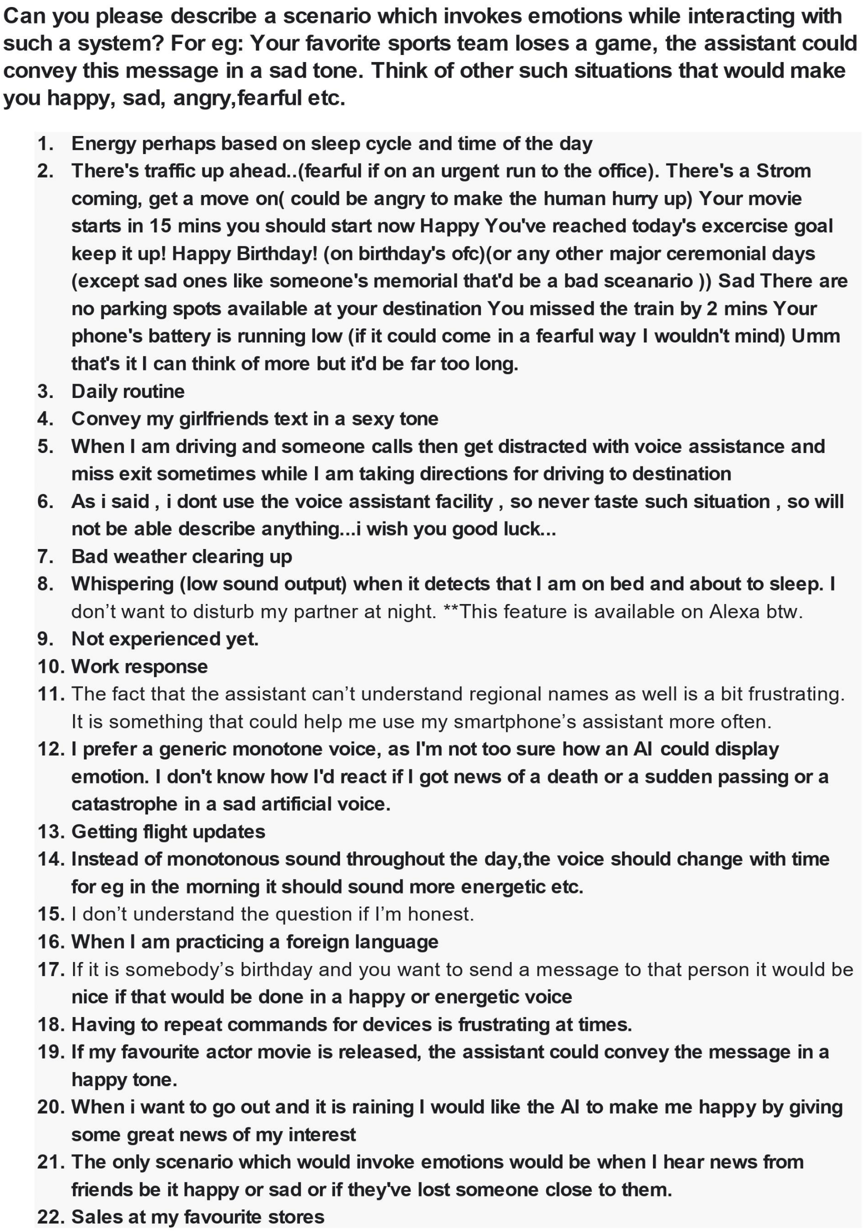
I accept the conditions described above and give approval for the use of my data in the form of text and recorded audio during this workshop for the purpose of this research. I understand that this data will be processed anonymously. *

□ Agree

□ Disagree

Figure A.2: Consent form for the focus groups

A.1.3 Online Survey to recruit participants

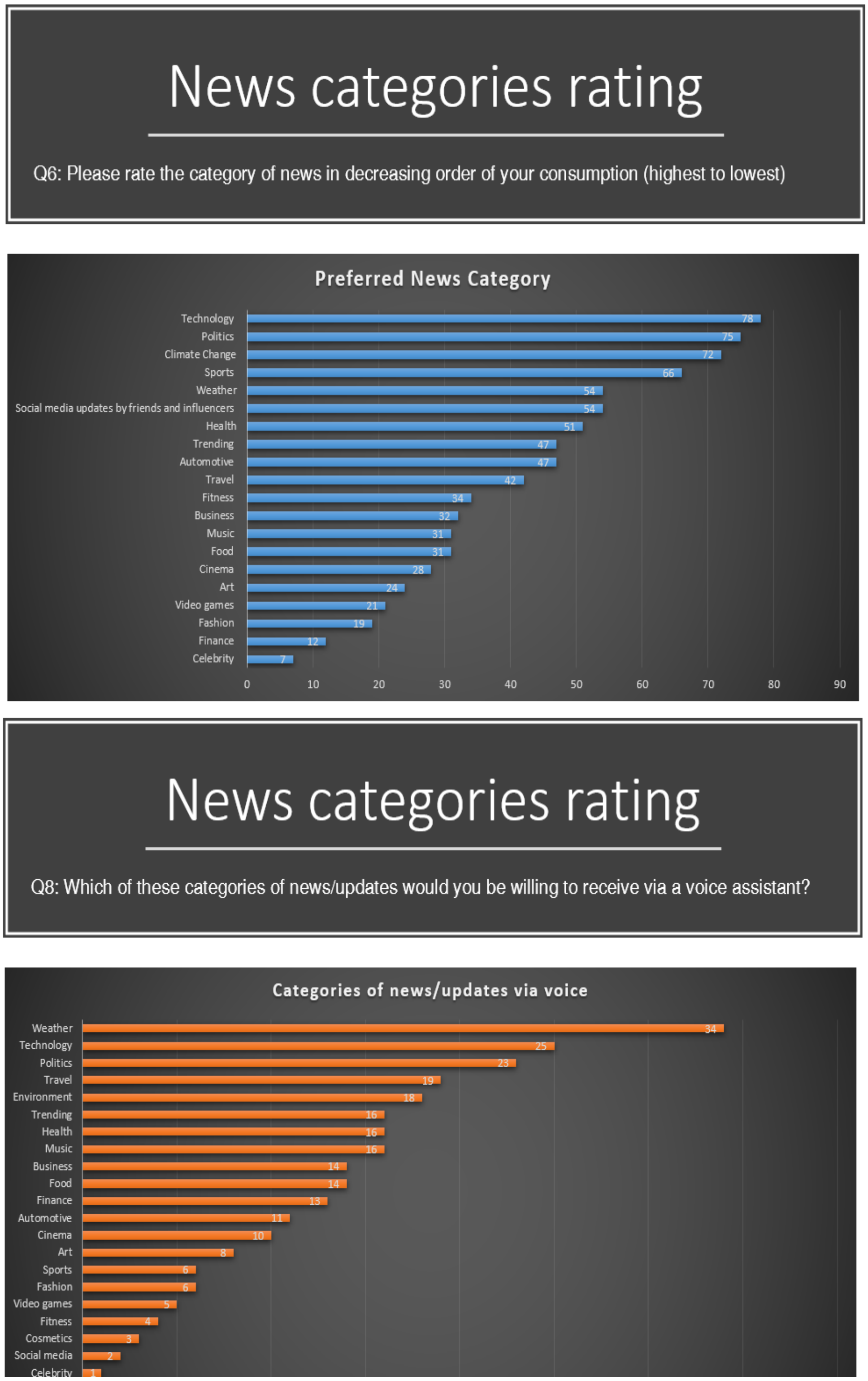


Can you please describe a scenario which invokes emotions while interacting with such a system? For eg: Your favorite sports team loses a game, the assistant could convey this message in a sad tone. Think of other such situations that would make you happy, sad, angry,fearful etc.

1. Energy perhaps based on sleep cycle and time of the day
2. There's traffic up ahead..(fearful if on an urgent run to the office). There's a Strom coming, get a move on(could be angry to make the human hurry up) Your movie starts in 15 mins you should start now Happy You've reached today's excercise goal keep it up! Happy Birthday! (on birthday's ofc)(or any other major ceremonial days (except sad ones like someone's memorial that'd be a bad sceanario)) Sad There are no parking spots available at your destination You missed the train by 2 mins Your phone's battery is running low (if it could come in a fearful way I wouldn't mind) Umm that's it I can think of more but it'd be far too long.
3. Daily routine
4. Convey my girlfriends text in a sexy tone
5. When I am driving and someone calls then get distracted with voice assistance and miss exit sometimes while I am taking directions for driving to destination
6. As i said , i dont use the voice assistant facility , so never taste such situation , so will not be able describe anything...i wish you good luck...
7. Bad weather clearing up
8. Whispering (low sound output) when it detects that I am on bed and about to sleep. I don’t want to disturb my partner at night. **This feature is available on Alexa btw.
9. Not experienced yet.
10. Work response
11. The fact that the assistant can’t understand regional names as well is a bit frustrating. It is something that could help me use my smartphone’s assistant more often.
12. I prefer a generic monotone voice, as I'm not too sure how an AI could display emotion. I don't know how I'd react if I got news of a death or a sudden passing or a catastrophe in a sad artificial voice.
13. Getting flight updates
14. Instead of monotonous sound throughout the day,the voice should change with time for eg in the morning it should sound more energetic etc.
15. I don’t understand the question if I’m honest.
16. When I am practicing a foreign language
17. If it is somebody’s birthday and you want to send a message to that person it would be nice if that would be done in a happy or energetic voice
18. Having to repeat commands for devices is frustrating at times.
19. If my favourite actor movie is released, the assistant could convey the message in a happy tone.
20. When i want to go out and it is raining I would like the AI to make me happy by giving some great news of my interest
21. The only scenario which would invoke emotions would be when I hear news from friends be it happy or sad or if they've lost someone close to them.
22. Sales at my favourite stores

Figure A.3: Responses for Survey to recruit focus group participants

A.1.4 Online Survey to prioritize use-cases



News categories rating

Q6: Please rate the category of news in decreasing order of your consumption (highest to lowest)

News categories rating

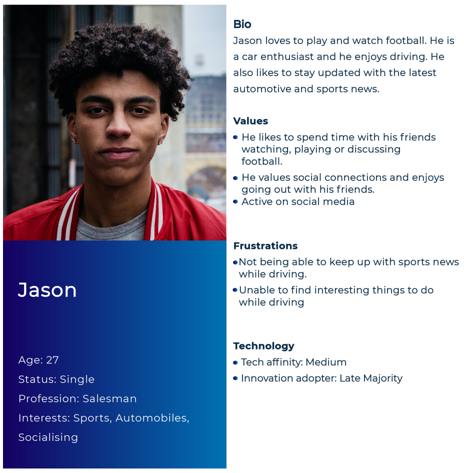
Q8: Which of these categories of news/updates would you be willing to receive via a voice assistant?

Figure A.4: Weighted Averages for use-case prioritization survey

A.1.5 List of Identified Use-cases

1	Dealing with an accident (e.g. Ask if you are safe, calm you down) or Accident nearby.
2	Stop sign warning detection
3	Suggest music and podcasts on long drives
4	Read out class notes or custom text (such as news or magazines)
5	Calm down drivers in case of road rage
6	Route options (take route highly rated or eco-friendly/ faster/ less traffic/ explorer route)
7	Suggest a highly rated service station or restaurant (cleanliness/taste/ value/ time)
8	Comforting voice when you miss a turn (give context on the turn, follow car in front)
9	Smile to start the car in the morning (good way to start your day)
10	Weather updates (rain snow stopping/ sun)
11	Emotions when reading out the message (according to the context)
12	Repeating commands is frustrating
13	Sing along, happy or sad drive
14	Seeing an accident (roadkill)
15	Curious about buildings or advertisements
16	Information about other cars/bikes that are around
17	Take a selfie feature when they start
18	Share my view (pictures of my drive)
19	Updates on share prices of stocks
20	Politics or breaking news
21	Driving to the gym (workout details/schedule)
22	Cheering up the driver then they have a bad day in the office (jokes or trivia)
23	Environmental news
24	Warnings on carbon footprint reaching close to user goals (sad)
25	Fun game to entertain the driver to avoid monotony
26	Suggest drop into the baby radio
27	Energetic in the morning, calmer in the evening
28	Inform the user about cancelled, postponed or rescheduled meetings
29	Save proactive notifications in my list (watch later playlist)
30	Announce updates or tutorials for new car functions
31	Social media updates of close friends or selected profiles
32	Asking the driver if they had alcohol (when they go to a bar or restaurant)
33	Parking assistant (good job when parked properly)
34	Comments on driving style (caution or appreciation)
35	Traffic updates (Sad for congestion up ahead/ happy for traffic clearing out)
36	Speed camera coming up, slow down or speed up
37	Keeping babies or kids engaged with games, whisper mode
38	Live updates for sports or sports news
39	Suggest Instagram post, story because they wore a new outfit
40	Suggests trends based on people's clothes in high street etc.
41	Warn about erratic behavior by other vehicles
42	New episode, movie or game is out
43	Smart-home notifications such as visitor alerts or delivery guy close by
44	Parking unavailable (sad) and opposite (happy)
45	Birthday or other celebrations (happy)

A.1.6 Personas



Jason

Age: 27
Status: Single
Profession: Salesman
Interests: Sports, Automobiles, Socialising

Bio

Jason loves to play and watch football. He is a car enthusiast and he enjoys driving. He also likes to stay updated with the latest automotive and sports news.

Values

- He likes to spend time with his friends watching, playing or discussing football.
- He values social connections and enjoys going out with his friends.
- Active on social media

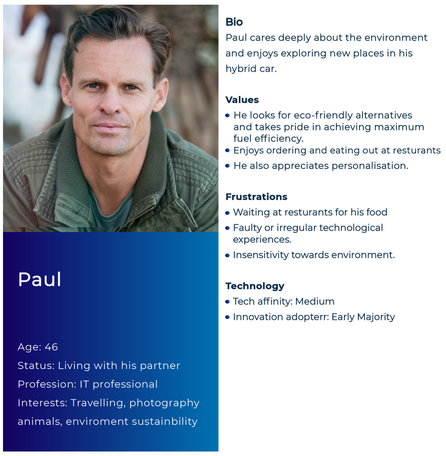
Frustrations

- Not being able to keep up with sports news while driving.
- Unable to find interesting things to do while driving

Technology

- Tech affinity: Medium
- Innovation adopter: Late Majority

Figure A.5: First Persona



Paul

Age: 46
Status: Living with his partner
Profession: IT professional
Interests: Travelling, photography animals, enviroment sustainbility

Bio

Paul cares deeply about the environment and enjoys exploring new places in his hybrid car.

Values

- He looks for eco-friendly alternatives and takes pride in achieving maximum fuel efficiency.
- Enjoys ordering and eating out at resturants
- He also appreciates personalisation.

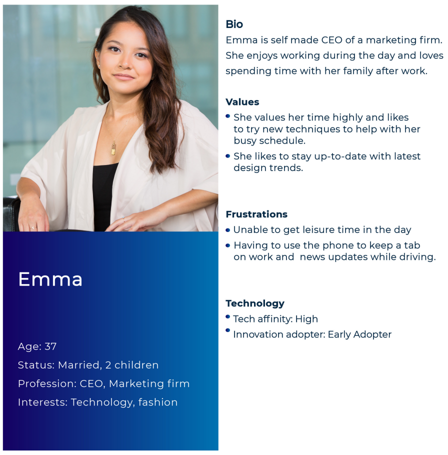
Frustrations

- Waiting at resturants for his food
- Faulty or irregular technological experiences.
- Insensitivity towards environment.

Technology

- Tech affinity: Medium
- Innovation adopterr: Early Majority

Figure A.6: Second Persona



Emma

Age: 37
Status: Married, 2 children
Profession: CEO, Marketing firm
Interests: Technology, fashion

Bio

Emma is self made CEO of a marketing firm. She enjoys working during the day and loves spending time with her family after work.

Values

- She values her time highly and likes to try new techniques to help with her busy schedule.
- She likes to stay up-to-date with latest design trends.

Frustrations

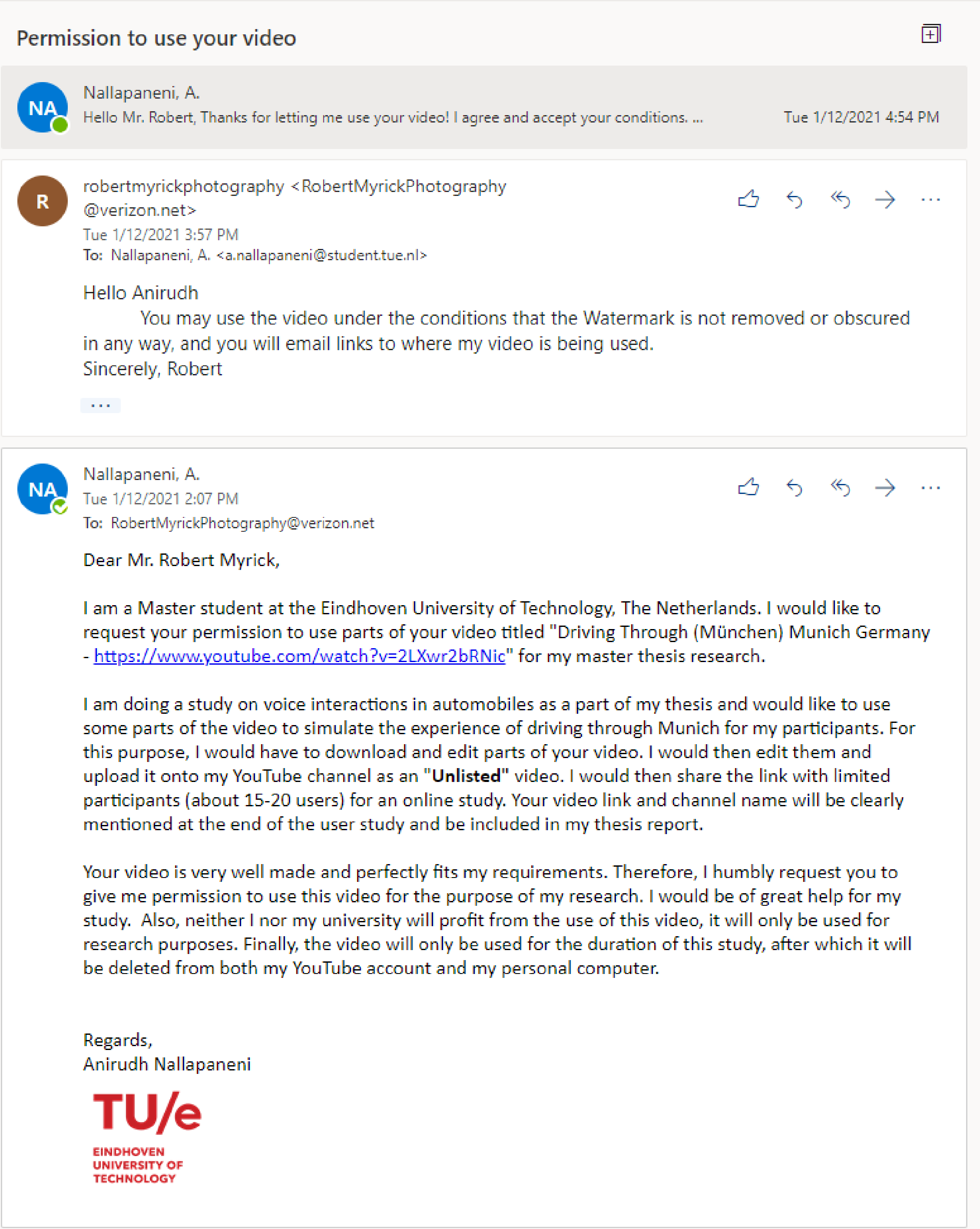
- Unable to get leisure time in the day
- Having to use the phone to keep a tab on work and news updates while driving.

Technology

- Tech affinity: High
- Innovation adopter: Early Adopter

Figure A.7: Third Persona

A.1.7 Consent to use video



Permission to use your video

Nallapaneni, A.
Hello Mr. Robert, Thanks for letting me use your video! I agree and accept your conditions. ...
Tue 1/12/2021 4:54 PM

robertmyrickphotography <RobertMyrickPhotography@verizon.net>
Tue 1/12/2021 3:57 PM
To: Nallapaneni, A. <a.nallapaneni@student.tue.nl>

Hello Anirudh

You may use the video under the conditions that the Watermark is not removed or obscured in any way, and you will email links to where my video is being used.

Sincerely, Robert

Nallapaneni, A.
Tue 1/12/2021 2:07 PM
To: RobertMyrickPhotography@verizon.net

Dear Mr. Robert Myrick,

I am a Master student at the Eindhoven University of Technology, The Netherlands. I would like to request your permission to use parts of your video titled "Driving Through (München) Munich Germany - https://www.youtube.com/watch?v=2LXwr2bRNic" for my master thesis research.

I am doing a study on voice interactions in automobiles as a part of my thesis and would like to use some parts of the video to simulate the experience of driving through Munich for my participants. For this purpose, I would have to download and edit parts of your video. I would then edit them and upload it onto my YouTube channel as an "**Unlisted**" video. I would then share the link with limited participants (about 15-20 users) for an online study. Your video link and channel name will be clearly mentioned at the end of the user study and be included in my thesis report.

Your video is very well made and perfectly fits my requirements. Therefore, I humbly request you to give me permission to use this video for the purpose of my research. I would be of great help for my study. Also, neither I nor my university will profit from the use of this video, it will only be used for research purposes. Finally, the video will only be used for the duration of this study, after which it will be deleted from both my YouTube account and my personal computer.

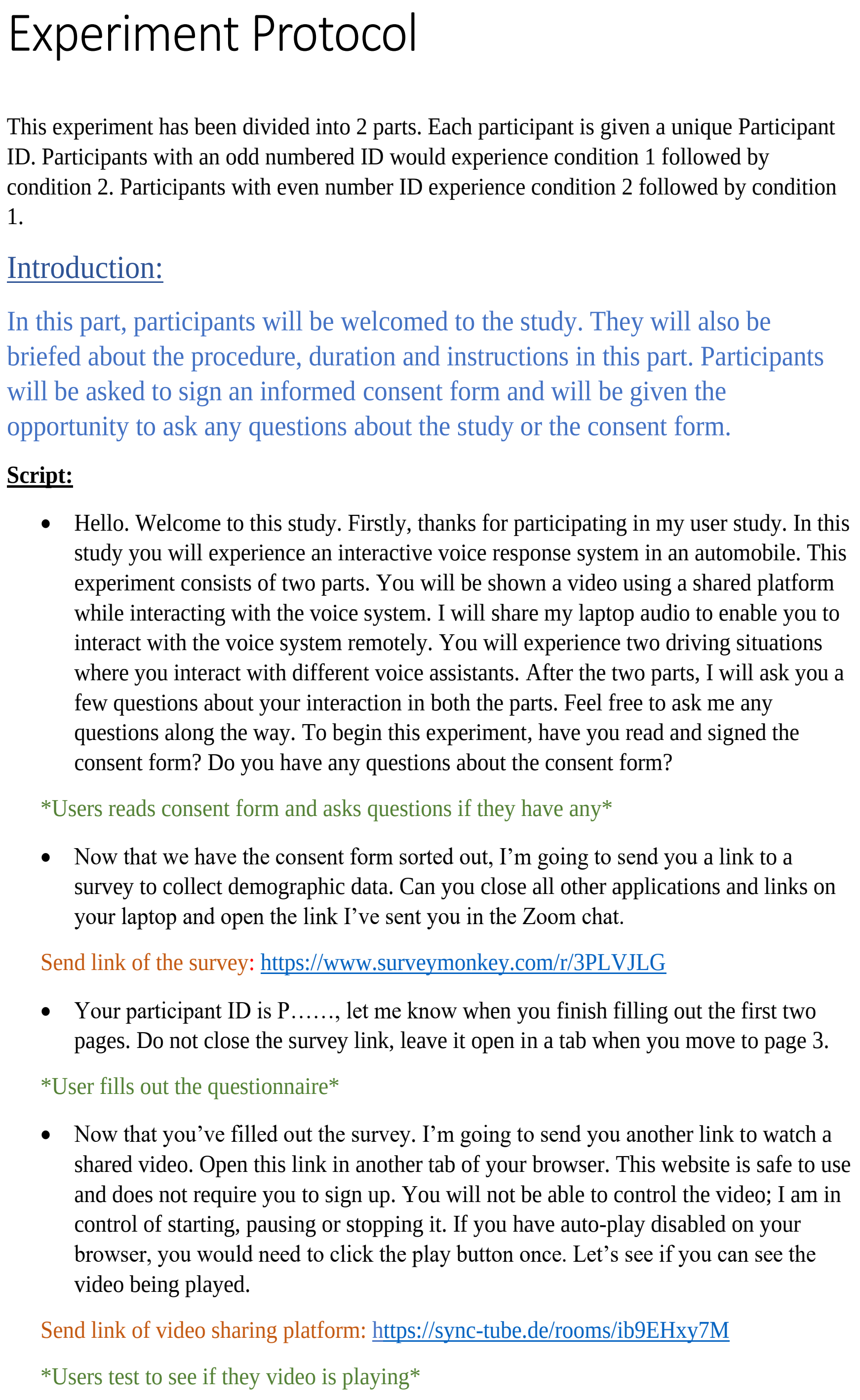
Regards,
Anirudh Nallapaneni

TU/e
EINDHOVEN
UNIVERSITY OF
TECHNOLOGY

Figure A.8: Permission to use YouTube video for study

A.2 Methodology

A.2.1 Experiment Script



Experiment Protocol

This experiment has been divided into 2 parts. Each participant is given a unique Participant ID. Participants with an odd numbered ID would experience condition 1 followed by condition 2. Participants with even number ID experience condition 2 followed by condition 1.

Introduction:

In this part, participants will be welcomed to the study. They will also be briefed about the procedure, duration and instructions in this part. Participants will be asked to sign an informed consent form and will be given the opportunity to ask any questions about the study or the consent form.

Script:

- Hello. Welcome to this study. Firstly, thanks for participating in my user study. In this study you will experience an interactive voice response system in an automobile. This experiment consists of two parts. You will be shown a video using a shared platform while interacting with the voice system. I will share my laptop audio to enable you to interact with the voice system remotely. You will experience two driving situations where you interact with different voice assistants. After the two parts, I will ask you a few questions about your interaction in both the parts. Feel free to ask me any questions along the way. To begin this experiment, have you read and signed the consent form? Do you have any questions about the consent form?

Users reads consent form and asks questions if they have any

- Now that we have the consent form sorted out, I'm going to send you a link to a survey to collect demographic data. Can you close all other applications and links on your laptop and open the link I've sent you in the Zoom chat.

Send link of the survey: https://www.surveymonkey.com/r/3PLVJLG

- Your participant ID is P……, let me know when you finish filling out the first two pages. Do not close the survey link, leave it open in a tab when you move to page 3.

User fills out the questionnaire

- Now that you've filled out the survey. I'm going to send you another link to watch a shared video. Open this link in another tab of your browser. This website is safe to use and does not require you to sign up. You will not be able to control the video; I am in control of starting, pausing or stopping it. If you have auto-play disabled on your browser, you would need to click the play button once. Let's see if you can see the video being played.

Send link of video sharing platform: https://sync-tube.de/rooms/ib9EHxy7M

Users test to see if they video is playing

- I will now share my computer audio so that you can interact with the IVR. The system has limited functionality, so try not to ask it complicated questions. A bit of a backstory for you, before we begin the first phase of the study. Imagine that you are in your car in your city. Also assume that the system knows about your likes and dislikes and has access to your calendar etc. For some context before we start the video, you missed last's night's game because you went out with a few friends. You will now be on a short ride in your car on a known route. Do not focus too much on the content of the video, it's just to put you into the simulated state of driving. We will ask you to evaluate the voice system with a few questions after this phase. Now to start the experiment, let's see if you can interact with the IVR.

Phase 1:

Starts playing phase 1 of the experiment in which the IVR introduces itself and asks the participant if they are ready to start driving.

- Now that you can hear the assistant, I am going to mute myself and start playing the video while you interact with the voice system. If you need me, just call out my name and I'll jump back into this call. Best of luck!

Cues to trigger each block in the video:

- **Weather:** Prompt after a U-turn at **1:06**
- **Sports:** Crossing the signal after the stadium at **1:52**
- **Trending:** Two pedestrians walk into the frame from the park at **2:40**
- **Parking:** Black VW Golf on the left at **3:55**

Users experience phase 1 of the experiment

- You've finished phase 1 of the experiment. Could you go back to the survey link and move to the next page? Let me know when you have finished answering the questions on this page and moved to page 4. Again, do not close the tab for the survey when you're done with page 3.

Users fill out the questionnaire on page 3

Phase 2:

- You will now experience phase 2 of this study. Let me start the prototype and see if you can interact with it.

Starts playing phase 2 of the experiment. The IVR introduces itself and asks the participants if they are ready to start driving.

- I am going to mute myself again and start playing the video, while you interact with the voice system. If you need me, just call out my name and I'll jump back into this call. Best of luck!

Cues to trigger each block in the video:

- **Weather:** Prompt after a U-turn at **1:06**

- **Sports:** Crossing the signal after the stadium at **1:52**
- **Trending:** Two pedestrians walk into the frame from the park at **2:40**
- **Parking:** Black VW Golf on the left at **3:55**

Users experience phase 2 of the experiment

- You have completed both the phases of the experiment. Now, I will start recording and ask you a few questions about the two phases. Do I have your permission to start recording?

Starts recording

Qualitative Interview: In this phase, a qualitative interview is conducted to receive subjective feedback about the voice prototypes.

The first question I have for you is:

1. Did you notice any differences between the two assistants? If so, what were they?
   Play sample of both voices to the participants if don't remember or cannot perceive a difference (works as manipulation check)

2. Which assistant do you prefer and why?

3. What do you think of voice interfaces showing emotions?

4. Did you find a difference between the positive and negative news from the VA in Condition 1? If so, which would you rate better?

5. Can you think of other situations where the voice interface can show emotions?

6. What did you think about the proactive nature of voice interface?

7. Rate the use cases by the order of your preference. The ones you experienced during this study were 1. Weather, 2. Sports, 3. Health/Trending, 4. Parking

8. What other contexts can a voice assistant be proactive?

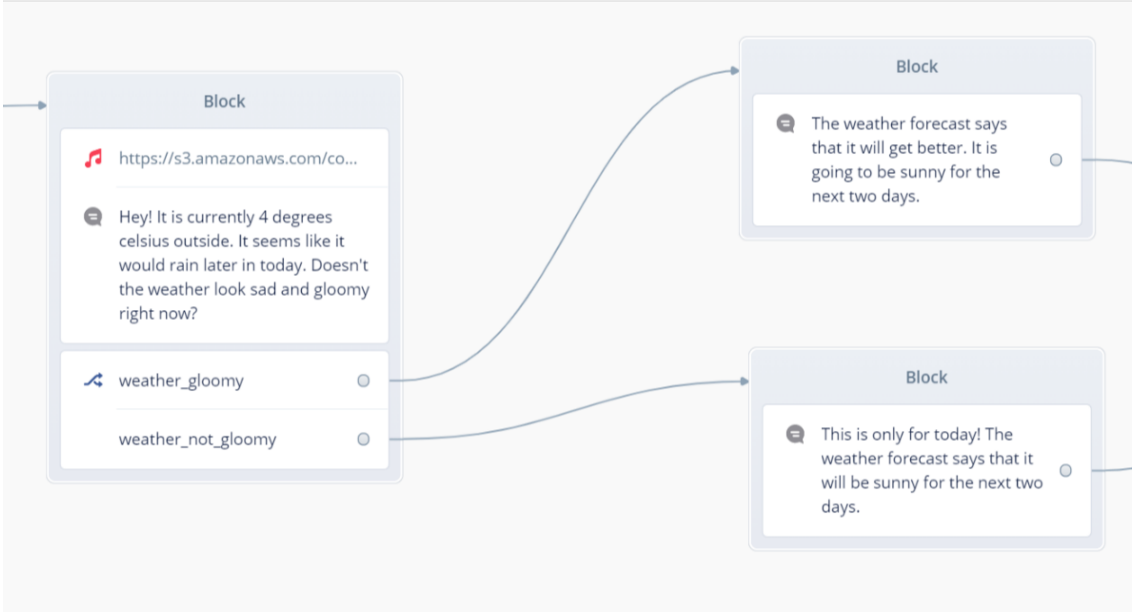
9. Would you be willing to use a system capable of showing emotions? Why or why not? Would you be willing to use a proactive system? Why/ why not?

10. That's it for the interview. Do you have any questions about the study or the voice prototype?

Users asks questions

Thank you for taking part in my study. Have a nice day! Bye!

A.2.2 Conversation Flow Control

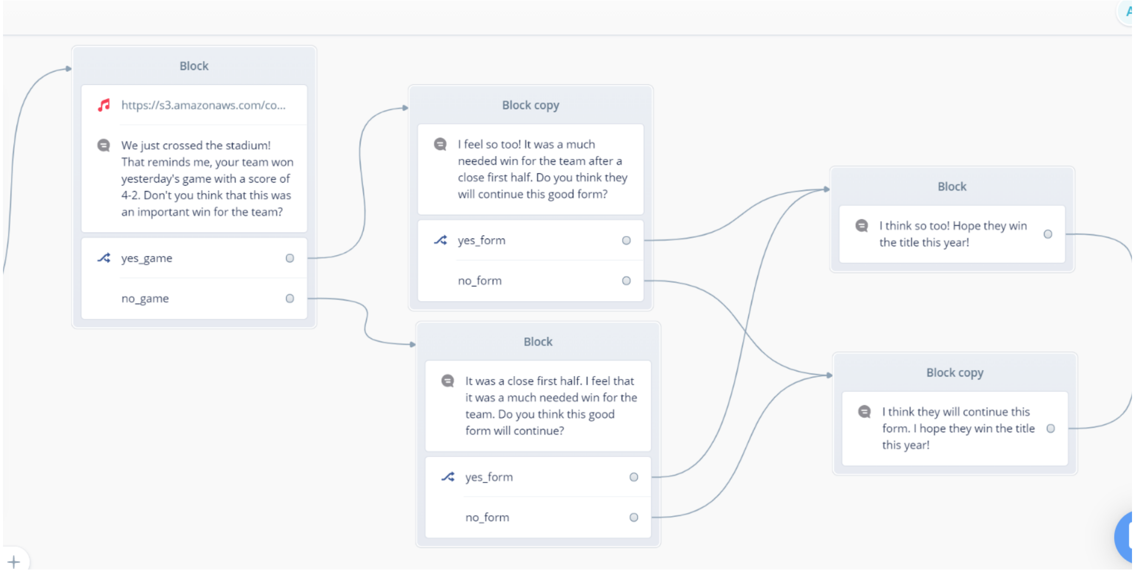


Hey! It is currently 4 degrees celsius outside. It seems like it would rain later in today. Doesn't the weather look sad and gloomy right now?

The weather forecast says that it will get better. It is going to be sunny for the next two days.

This is only for today! The weather forecast says that it will be sunny for the next two days.

Figure A.9: Non emotional weather conversation flow



We just crossed the stadium! That reminds me, your team won yesterday's game with a score of 4-2. Don't you think that this was an important win for the team?

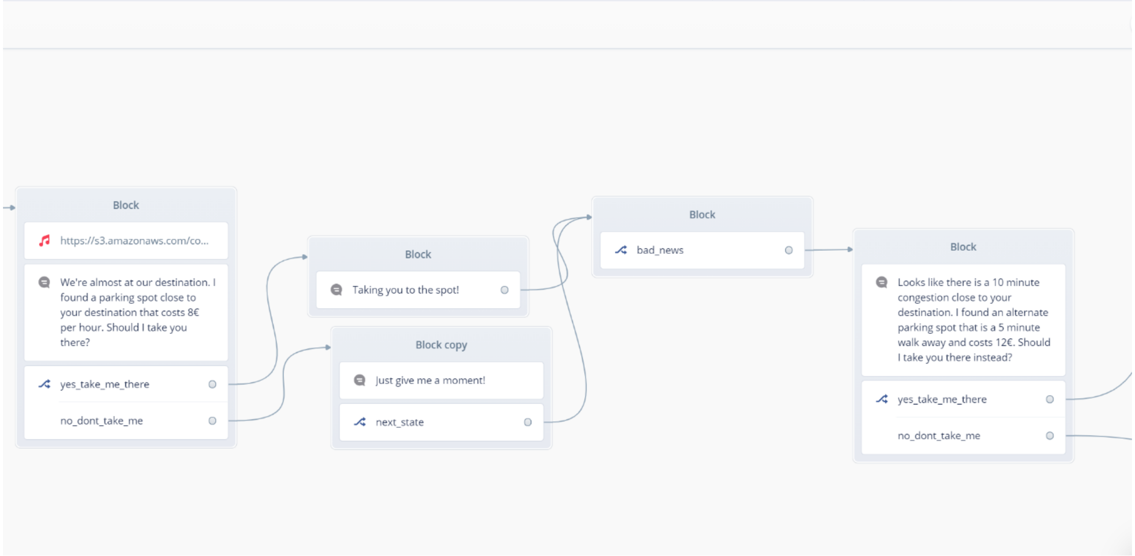
I feel so too! It was a much needed win for the team after a close first half. Do you think they will continue this good form?

It was a close first half. I feel that it was a much needed win for the team. Do you think this good form will continue?

I think so too! Hope they win the title this year!

I think they will continue this form. I hope they win the title this year!

Figure A.10: Non emotional health conversation flow



We're almost at our destination. I found a parking spot close to your destination that costs 8€ per hour. Should I take you there?

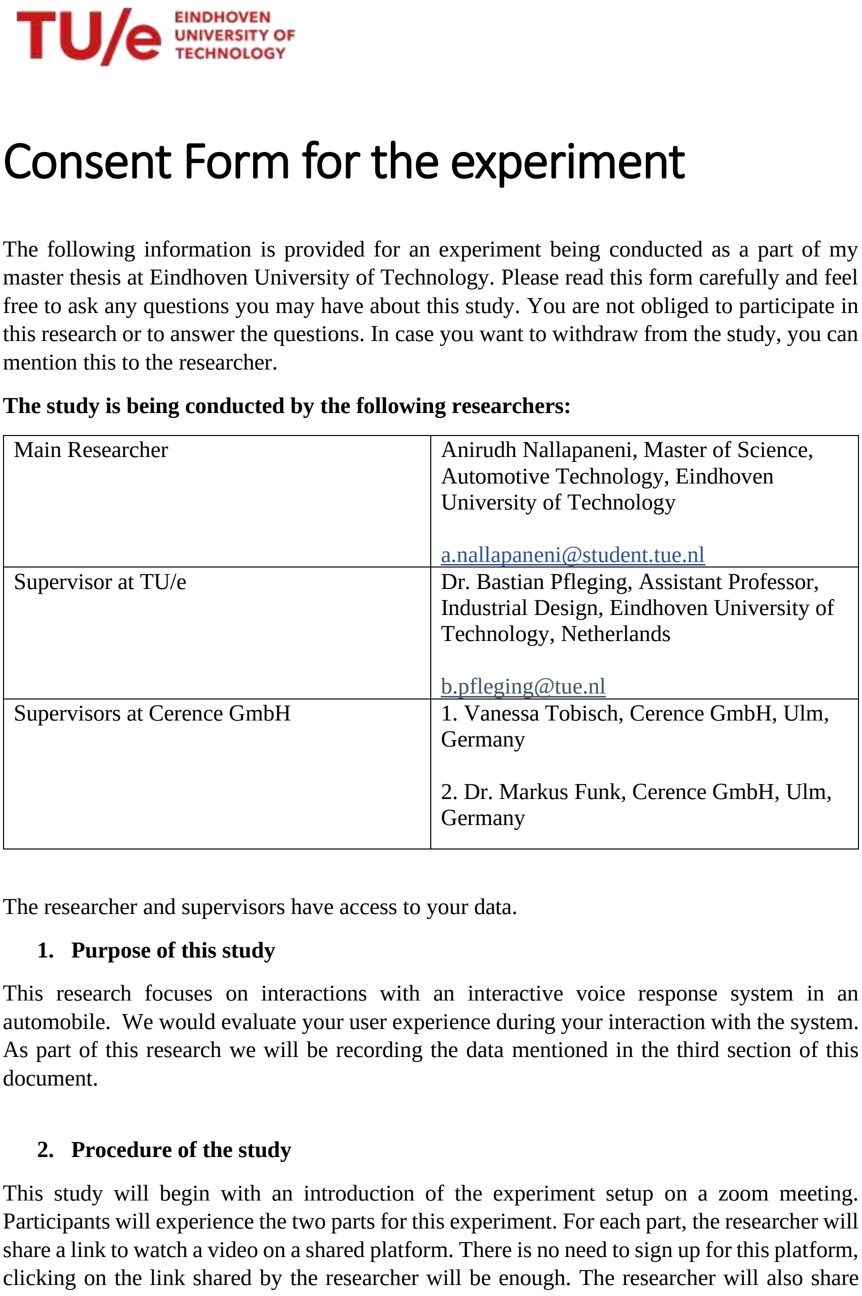
Taking you to the spot!

Just give me a moment!

Looks like there is a 10 minute congestion close to your destination. I found an alternate parking spot that is a 5 minute walk away and costs 12€. Should I take you there instead?

Figure A.11: Non emotional parking conversation flow

A.2.3 Consent form for the experiment



Consent Form for the experiment

The following information is provided for an experiment being conducted as a part of my master thesis at Eindhoven University of Technology. Please read this form carefully and feel free to ask any questions you may have about this study. You are not obliged to participate in this research or to answer the questions. In case you want to withdraw from the study, you can mention this to the researcher.

The study is being conducted by the following researchers:

Main Researcher	Anirudh Nallapaneni, Master of Science, Automotive Technology, Eindhoven University of Technology a.nallapaneni@student.tue.nl
Supervisor at TU/e	Dr. Bastian Pfleging, Assistant Professor, Industrial Design, Eindhoven University of Technology, Netherlands b.pfleging@tue.nl
Supervisors at Cerence GmbH	1. Vanessa Tobisch, Cerence GmbH, Ulm, Germany 2. Dr. Markus Funk, Cerence GmbH, Ulm, Germany

The researcher and supervisors have access to your data.

1. **Purpose of this study**

This research focuses on interactions with an interactive voice response system in an automobile. We would evaluate your user experience during your interaction with the system. As part of this research we will be recording the data mentioned in the third section of this document.

2. **Procedure of the study**

This study will begin with an introduction of the experiment setup on a zoom meeting. Participants will experience the two parts for this experiment. For each part, the researcher will share a link to watch a video on a shared platform. There is no need to sign up for this platform, clicking on the link shared by the researcher will be enough. The researcher will also share their computer audio via Zoom which would enable participants to interact with the voice assistant. The duration of each block will be about 12-15 minutes. After each phase, participants will have a fill a questionnaire. After the second phase participants will be asked a few questions about their interaction with the system.

The total experiment duration is about **30 - 35 minutes.**

3. **Description of potential discomforts, risks or inconveniences of participating in this study**

There are no foreseeable risks, discomforts with regards to participating in this user study. Participants will interact with the prototype by purely using their voice. The questionnaire and interviews during this experiment includes only low risk information with the results being presented in an aggregated form. A slight inconvenience such as simulator sickness could occur due the visuals shown to participants. However, the visuals are in the form of videos presented, which have a low risk of inducing simulator/motion sickness. Participants will not be exploited, and none of their data would place participants at risk of criminal or civil liability.

4. **Compensation**

As this project is not funded, we cannot provide any compensation for taking part in the study.

5. **Data to be collected during this study**

We will be recording the following data for the purpose of this study. No personal data such as your name, email address and phone number will be collected. You will be given a **Unique Participant ID** which cannot be traced back to you.

The following data will be collected during this study

I. Demographic data
   a. Gender
   b. Age Range
   c. Country of residence
   d. Driving Experience

II. Questionnaire Responses

III. Audio recordings of the qualitative interviews after each phase.

Audio recordings will **only** be used for transcribing your answers in the interview.

6. **Anticipated knowledge gain**

Through the collected data we want to gain insights into how voice interface concepts in automobiles impact the user experience of participants.

7. **Data protection and storage**

The data collected in this study will be coded and assigned to a random number. The coded data cannot be traced back to you. It will be stored locally on the password protected devices of the researchers. Data that is analysed will be stored on a secure password protected institutional repository at Eindhoven University of Technology or on a provide of such a service (e.g. SurfDrive/ResearchDrive). It will be stored in accordance with the Netherlands

Code of Conduct for Research Integrity i.e. 10 years. When shared with the partner, only anonymized data will be shared, through secure platforms that adhere to the data protection standards set by the TU/e as well as GDPR (such as SurfDrive or ResearchDrive).

8. Legal Basis

The legal basis for the processing of your personal data in scientific studies is your voluntary written consent according to EU GDPR. You have the following rights about your data (Article 13 et seq. EU GDPR):
Right to information: You have the right to be informed about the personal data concerning you which is collected, processed or, if applicable, transferred to third parties within the framework of the scientific study (provision of a free copy) (Article 15 EU GDPR).

Right to rectification: You have the right to have incorrect personal data concerning you corrected (Articles 16 and 19 EU GDPR).
Right to deletion: You have the right to delete personal data concerning you, e.g. if these data are no longer necessary for the purpose for which they were collected (Articles 17 and 19 EU GDPR).
Right to limitation of processing: Under certain circumstances, you have the right to request that the processing be restricted, i.e. the data may only be stored, not processed. You must request this. For this purpose, please contact your auditor or the data protection officer of the test centre (Articles 18 and 19 EU GDPR).
Right to data transferability: They have the right to obtain the personal data concerning them that they have provided to the person responsible for the clinical trial. This will enable you to request that this data be communicated either to you or, where technically possible, to another body designated by you (Article 20 EU GDPR).
Right of objection: You have the right to object at any time to concrete decisions or measures regarding the processing of your personal data (Art 21 EU GDPR). Such processing will no longer take place afterwards.

9. Data confidentiality

All personal data collected during the study will be processed confidentially and test subjects will never be made recognizable in publications, academic material or any other mean. For illustrations in publications and other educational/academic material, the researchers will substitute the participants in pictures / videos. All people involved in the experiment will be instructed on the importance of data privacy and security, to maintain confidentiality, and are required to follow procedures as outlined for instance in the EU GDPR regulations.

Finally, no individual results will be published, as conclusions will be made from the entire cohort's data. The results of this study will be disseminated in scientific conferences and published in conference proceedings, scientific research journals, project reports, student theses, and standard press and social media (advertising the actual research papers). Participants can ask the researchers for an electronic copy of the data that he/she has provided or that have been measured directly. If they are dissatisfied with how data privacy is handled, they can submit a complaint to the Chief Information & Security Officer, the Privacy & Security Officer and/or the Data Protection Officer of the Eindhoven University of Technology via privacy@tue.nl or contact the Dutch Data Protection Authority.

10. Withdrawing Consent

You are not obligated to participate in this research or to answer the questions, and you are also free to withdraw from the survey at any time. This applies to the data collected during study, its storage and use for future research. The study data collected until the moment you withdraw your consent will still be used in the study.

Consent to the processing of personal data and right to revoke this consent

The processing of your personal data is only lawful with your consent (Article 6 EU GDPR). You have the right to revoke your consent to the processing of personal data at any time. However, the data collected up to this point may be processed by the bodies named in the study information and declaration of consent for the respective scientific study (Article 7, paragraph 3 EU GDPR). If you wish to make use of one of these rights, please contact your examiner or the data protection officer at your test center.

Consent

I was informed in writing about the nature, significance, implications and risks of the scientific study and had ample opportunity to clarify my questions in an interview with the researcher.

☐ **Yes.** By confirming**,** you give permission to use your inputs during this session and use your data for the purpose of this research. You understand that this data will be stored and processed anonymously. You also consent for your data to be used for follow-up research.

☐ **No.** By disagreeing, you do not give permission to use your inputs during this session and use your data for the purpose of this research or follow-up research.

Signature:

Name –

Participant ID –

Place –

Date –

The above-mentioned personal data is for the purpose of identifying your consent. It will be kept separate from the data collected during the study. Your data will be assigned to a Unique Participant ID, which will be anonymized and decoupled from the data presented above, so that it cannot be traced back to you.

Contact Details of the researcher: **Anirudh Nallapaneni -** a.nallapaneni@student.tue.nl