Music Control in the Car –
Designing voice interactions
between user and a music
service with a focus on in-car
usage

CAROLINE ARKENSON
Kontroll av Musik i Bilen – Design av röstinteraktion mellan användare och musiktjänst med fokus på användning i bilen

Caroline Arkenson

SAMMANFATTNING

Musik är en viktig del i människors liv och tack vare teknikutveckling är musik mer tillgängligt än någonsin. Det skapar en behaglig och musikalisk miljö i flertalet situationer såsom under bilkörning, och musikströmming i bilen är en av de mest efterfrågade funktionerna bland bilförare. Användning av teknik såsom mobila enheter under bilkörning är dock en viktig säkerhetsfråga vilken kan åtgärdas genom att implementera röststyrning då en lyckad implementation av röststyrning kräver lägre kognitivt fokus av bilföraren än användandet av manuella kontroller eller att titta på en skärm.

Detta examensarbete har resulterat i ett antal designförslag för röstinteraktion mellan bilförare och en musiktjänst för en säker och trevlig användarupplevelse, där ett mål är att bilföraren via dialogen ska få en känsla av bekantskap med musiktjänsten och dess varumärke. Designförslagen är baserade på riktlinjer för design samt tidigare forskning inom röstinteraktion och teknik i bilen, då även inkluderat kognitiv belastning på föraren. Åsikter från bilförare och musiklyssnare har samlats in via en enkät om att kontrollera musikapplikationer under bilkörning och en designworkshop där deltagarna, baserat på olika historier, designade dialoger mellan bilförare och musiktjänsten.

Studien fann att bilförare är väldigt säkerhetsmedvetna och att de vill på ett säkert och snabbt söka efter specifik musik eller musik som passar det nuvarande humöret. När det kommer till igenkänningsfaktorn bör bilförare vid röststyrning kunna använda existerande funktioner i musiktjänsten och kunna använda funktioner som påverkar ens egna användarkonto - såsom att kunna spara låtar i en spellista.
Music control in the car – designing voice interactions between user and a music service with a focus on in-car usage

Caroline Arkenson
KTH Royal Institute of Technology
Stockholm, Sweden
arkenson@kth.se

ABSTRACT

Music is an important part of people’s lives and thanks to technology development music is more accessible than ever. It creates a pleasant auditory environment in many situations such as when driving, and in-car music streaming is one of the most requested features amongst drivers. Using technology such as mobile devices when driving is a major safety concern which can be addressed by implementing voice control. Said modality, if implemented successfully, puts less cognitive demand on the driver than using manual controls or looking at a monitor.

This thesis has resulted in a set of design suggestions for voice interaction between driver and a music service for a safe and pleasant user experience where the dialogues aim at providing drivers with a feeling of familiarity and brand recognition with the music service. The design suggestions are based on design guidelines and previous research on voice interaction and in-car technology including cognitive demands on the driver. Input from drivers and music listeners have been collected first through a questionnaire on controlling music applications while driving followed by a design workshop where the participants, based on different stories, designed dialogues between driver and a music service.

The research found that drivers are highly concerned with traffic safety, and want to be able to quickly and safely search for specific music or music that suits a current mood. In terms of familiarity, drivers should be able to use existing features from the music service when interacting through voice and be able to make actions that affect their private account – such as saving songs to a playlist.

1. INTRODUCTION

Music has been part of all human cultures dating far back in history and listening to music is one of the most rewarding experiences for humans (Salimpoor, et al., 2009). Consistently being ranked as one of the top ten things that people find highly pleasurable, music is a ubiquitous and important part of most people’s lives.

People listen to music for various reasons amongst some are mood enhancement, finding comfort when dealing with personal problems and defining one’s personal as well as social identity. With its possibility to create a pleasant auditory environment, music motivates people when performing boring tasks and often acts as an auditory decor in social gatherings. It is evident that music is a positive factor in our lives and thanks to technological development such as portable audio, digitalized music and music transferred over the Internet – i.e. streaming – music has become more accessible than ever. It is widespread and under individual control (Ter Bogt, et al., 2010). As technology has advanced, music has become the soundtrack of our everyday lives. It accompanies us in many day-to-day activities such as cooking, cleaning and driving (Frith, 2002).

1.1 Music in the car

Listening to music while driving is a common activity as the car is one of the environments where advancing technology has made access to music easier. From the installation of radio in the car to the possibility to connect various technological devices to the in-car music system, the opportunities for selecting what to listen to in the car has obviously increased with time. One thing that is clear is that people regulate the music to fit their current driving situation. For example, many drivers turn down the volume before performing a complex manoeuvre and on the contrary increase the volume when stuck in traffic jam (Dibben and Williamson, 2007).

A study by BI Intelligence on the connected car market reports that when it comes to implementing new features in the car 69% of consumers want in-car music streaming, making it the most desirable in-car feature of that study (Greenough, 2015). In other words, there is no lack of interest for music listening in the car. However, introducing technology and mobile applications for in-car usage raises a concern of safety. Driving is a very complex activity as it requires significant visual and cognitive attention in order for it to be performed successfully (Cooper, Ingebritsen and Strayer, 2014).

In-car technology can either be embedded or rely on secondary devices (Greenough, 2010). As mobile devices have become ubiquitous in everyday life, the use of such devices while driving has become a crucial traffic safety concern. Using a mobile device while driving causes cognitive interference with that of the driving task as it impairs the allocation of visual attention, impairs speed control and increases reaction time (Zhao, et al., 2013). As a reaction to that, drivers have expressed their wish to reduce the risks of using mobile devices while driving by being able to control their devices through voice interaction (Phillips, Nguyen and Mischke, 2010). Voice interaction allows the driver to control their devices through speech, and consequently the driver can keep their eyes on the road instead of their smartphones.

1.2 Goal of the thesis study

It is evident that music is a very important part of people’s lives and that music listeners want to be able to choose what to listen to. Music listeners are in control of the music they are listening to and as technology has improved music has become more accessible than ever – the car setting is no exception. Since using technology in a driving situation is a potential safety threat, voice interaction can help reducing the risk of paying attention on mobile devices when it should be on the road ahead.
The goal of this study – which was conducted in collaboration with Spotify – was to design suggestions for a set of voice interactions that improve a person's ability to safely interact with a music service while driving, and increase overall sentiment towards the music service and its brand. The way to reach the thesis goal was to, through research, answer the question “What voice interactions should be designed in order to provide hands-free control of an already existing music service, making sure that the dialogues provide users with a feeling of familiarity with the music service?”

The research question was divided into two sub-questions which led the research forward:

- What actions do users wish to perform while listening to music when driving?
- How should the music service communicate with its users?

1.3 Delimitations
The focus of this study was to design voice interactions, more specifically spoken input and output, with a focus on in-car usage. Due to time constraints and the nature of a thesis study, some areas relating to the field of study were not addressed.

Visual and auditory design was not considered to be part of the design task as audio and visual design are two fields separated from voice interaction design. As this study was conducted in English, the design suggestions were also presented in English. Localization – translations of voice interactions – is an extension that was not addressed.

2. BACKGROUND
The background section presents both positive aspects of and challenges with voice as a modality and introduces the reader to voice user interfaces and in-car technology. Included are cognitive demands on the driver as well as design guidelines for in-car technology and voice user interfaces.

An increasing challenge in human-computer interaction, due to higher demands on faster interaction between users and information, is making information access efficient and easy (Farinazzo, et al., 2010). In eyes-busy and/or hands-busy situations when users cannot access keyboards and/or monitors, such as when driving, the task of accessing information is obstructed (Nielsen, 2003). An expectation on information to always be accessible characterizes today’s mobile society and being highly multifunctional, mobile devices have become people’s primary personal communication devices (Meisel, 2010). While overall usability of mobile devices has improved with time, tasks like typing search queries can be cumbersome, error-prone and in some scenarios – like in the car – even be unsafe (Schalkwyk, et al., 2010).

2.1 Voice as a modality
Even though mobile devices are limited in terms of input and output due to screen size and the use of hands, they do provide several other modalities for information access. Voice is one of the modalities which mobile devices support, and voice interaction is a good way to enter and retrieve information when a system otherwise fails to support said tasks (Farinazzo, et al., 2010). Talking is after all the most natural form of communication for people and people can talk much faster than they can write (Phillips, Nguyen and Mischke, 2010). Additionally, searching by voice is not a new activity – we verbally ask questions to people in our surroundings and phone call services have been available for decades (Schalkwyk, et al., 2010).

Voice interaction has become more common in everyday life during the past years with applications like Google’s Google Now, Apple’s Siri and Amazon’s Alexa. Amazon Echo is a speaker equipped with microphones and provides information, answers questions and plays media by connecting to Alexa – a cloud-based voice service (Amazon, n.d.c). With Google Now and Siri, users can search for information, start applications on their mobile devices and schedule meetings (Google Inc., 2015; Apple Inc., n.d.). One thing that the three of these services have in common is that they can all be triggered by a “wake word” – Hey Siri, Ok Google and Alexa. By saying the wake word, the user lets the application know that a command should be recorded (Google Inc., 2015; Amazon, n.d.c.; Apple Inc., n.d).

2.1.2 Challenges with voice interaction
While voice interaction is suitable for certain scenarios and has improved during the last decades, implementing voice interaction is not without its challenges. Vocabularies are huge and input is unpredictable (Schalkwyk, et al., 2010) which means that there is no guarantee of correct interpretation of commands (Dahlgren, 2013). The average recognition rate is slightly above 90%, which can be frustrating for the user since it basically means that the system will misinterpret the user every tenth command (Klein, 2015) and there are various factors that affect the recognition rate.

Noise conditions may vary between development, training and different usage environments and could thus contribute to misinterpretation by the system. Mismatch could also occur in voice and phrases between training and actual use (Schalkwyk, et al., 2010; Dahlgren, 2013). An example of such a mismatch is paraphrasing of commands. The user might not remember exactly what the defined phrase was from training when uttering it for real use. In the example of a GPS, a phrase defined to be “Navigate to” might be rephrased by the driver as “Take me to” (Schalkwyk, et al., 2010). Dialectal variations (Schalkwyk, et al., 2010), pronunciation differences and everyday variability in speech affected by health or emotional state (Dahlgren, 2013) are also possible aspects that could affect the recognition.

Since the user is talking to the system, interrupting it when talking can feel unnatural and impolite to the user (Klein, 2015). Interrupting the system might also cause the user to get stuck in a frustrating loop of “I did not understand you”. If the user interrupts the system while it is talking, it could be difficult for the system to record the entire command as it is focusing on output rather than input at that moment.

2.2 Designing voice user interfaces
A voice user interface (VUI) is what users interact with when communicating with a spoken dialogue system. Prompts, grammars and dialogue logic are the components that build up a VUI. Prompts are pre-recorded system messages which ask the user for input, grammars are defined phrases which the user can say in response to each prompt and the dialogue logic defines actions taken by the system (Cohen, Giangola and Balogh, 2004).
Design principles for VUIs overlap with those of other types of interfaces, however, the characteristics of a VUI pose both unique design challenges and opportunities as the interaction is through spoken language. (Cohen, Giangola and Balogh, 2004). Nielsen (2003) argues that voice interfaces work best as an additional component of a multimodal system rather than as a standalone dialogue with the user, as a VUI is a medium with zero persistence. Due to its sequential nature, speech as the only output is very limited and allocates much of a user’s concentration. By making the system multimodal, it opens up the possibilities of taking advantage of the best aspects of both voice and visual modalities (Schalkwyk, et al., 2010). People can read much faster than they can listen (Phillips, Nguyen and Mischke, 2010), and effectively combining a screen with a VUI can significantly reduce the cognitive demands of a user and a multimodal approach broadens design constraints that come with designing a voice-only system (Cohen, Giangola and Balogh, 2004).

As with any kind of interaction design, what to say is a key issue and the main usability determinant. VUI designers need to decide the structure of dialogues between the user and the system, what tasks to support and available commands and features. Users should be able to specify what they want to achieve and the system needs to provide the user with adequate feedback (Nielsen, 2003). The user should feel in control of the interaction and the system clearly needs to indicate that it understood what the user said and that the commands are being processed (Farinazzo, et al., 2010).

VUIs can be thought of as a conversation between user and system, and as in a conversation with a human – we expect the responding party to have a personality. Successful voice interfaces have interesting personas as it will help designing better, consistent dialogues and steer design decisions regarding system response (Klein, 2015). The way the system talks should reflect the company’s brand and stay true to the application domain – an educational system should speak in a different way than a psychotherapy system. By giving the system a personality, users become more engaged which increases likeability, acceptance and user satisfaction (López-Cózar, et al., 2014). Choice of voice also affects the user’s perception of the VUI as people apply their gender stereotypes on computer and machine voices (Jeong and Shin, 2015).

Another element which greatly improves the usability of a VUI is context awareness (Nielsen, 2003). The most common definition of context is information relevant to the interaction between user and application that can be used to characterize the situation of an entity (López-Cózar and Callejas, 2010) and the user is considered to be part of the contextual information. The more a voice system knows about its surrounding environment and the current context, the greater the usability (Nielsen, 2003). Processing context is essential for dialogue systems aiming at working with the use of natural language (López-Cózar and Callejas, 2010). Natural language processing and establishing a suitable dialogue initiative between the system and user are key elements for VUI usability. So are vocabulary size and domain coverage (Farinazzo, et al., 2010).

2.2.1 Designing the conversation

Absolute semantic commands require the user to explicitly tell the system what they want to achieve (Dahlgren, 2013). These commands are very detailed and specific and a system relying solely on absolute semantic commands can only make decisions based on the words that the user said. Commands uttered by the user that are not detailed enough or missing a required word become error-prone. Context-dependent commands can refer to the user’s and application domain’s context. A system which understands context does not require as detailed commands. “Yes” and “No” are two commands that have different meaning depending on the context, but are sufficient for a context-aware system to understand what action to perform next. Systems using absolute semantic commands on the other hand have no knowledge of previous interactions with the user (Dahlgren, 2013).

With context-dependent commands, users are allowed to be less specific as the system relies on the current context to fill in a missing gap with information from the application domain (Dahlgren, 2013). Interacting using context-dependent commands is more natural to users as the commands can include less information than with absolute semantic commands. However, making use of both absolute semantic and context-dependent commands can decrease misinterpretation. If the user is in a conversation with another person, the conversation creates a context which the system should not be triggered by if not intended (Dahlgren, 2013).

In order to create a successful and satisfying user experience, it is suggested in the Alexa Skills Kit Voice Design Handbook (Amazon, n.d.b) that commands uttered by users should either be in form of a question or a statement. In case the user does not provide all parts needed for the system to fulfill a task, the system should ask for further information (Amazon, n.d.b).

It is suggested in The Alexa Skills Kit Voice Design Best Practices (Amazon, n.d.a) that the system clearly should present the user with different options to choose from or possible actions to take, and indicate when it prompts the user for a response in a dialogue. In case re-prompt is needed to fulfill a task, one piece of information should be requested at a time. Output should be kept short as the user needs to process it by listening, questions should only be asked if necessary and the user should not be overwhelmed with too many choices – not more than three options. The system should let the user know what the current application context is and in case of errors, offer a way for the user to get out of the error (Amazon, n.d.a).

2.3 In-car voice-based technology and cognitive demands

Visual distraction has a larger impact on driving performance and accident hazard than what cognitive distraction has (Peissner, Doebler and Metze, 2011). Therefore, voice interaction has been proven to be superior than manual control and graphical displays in terms of safe in-vehicle technology. It is, however, important to keep in mind that the design of a voice-based system greatly affects both the safety of the system and driver distraction.

In order to minimize distraction, reduce driving errors and not cause cognitive overload for drivers, various design principles are available to consider when designing systems for in-car usage (ustwo, 2014). The European statement of Principles on In-Vehicle Information and Communication Systems (European Commission, 1998) states that the driver’s attention to the system should be compatible with the attention needed in the driving situation; the system is not allowed to be designed in a way that
distracts the driver. It is the driver who should be in charge of the pace of interactions and the driver should be able to keep at least one hand on the steering wheel while controlling the system. Voice based systems should allow for hands-free interaction and as with any type of system usability, appropriate and clear feedback after input is required (European Commission, 1998). Voice as a modality in a driver’s situation is successful when it aims at increasing safety by minimizing distractions that come from manually controlling technologies in the car (ustwo, 2014).

It is difficult to effectively assess cognitive distraction from voice interaction both due to difficulties in observing the focus of a driver’s brain (Strayer, et al., 2014) and varying level of difficulty to drive different cars (Cooper, Ingebretsen and Strayer, 2014). Consequently, studies in the subject have generated mixed results regarding level of cognitive distraction (He, et al., 2015). It has been suggested that performing voice tasks in general is more demanding than tasks like engaging in conversations and listening to music. On a positive note, if the commands can be completed with minimal amount of errors in few steps the cognitive demand only increases somewhat (Cooper, Ingebretsen and Strayer, 2014) while at the same time improving driver performance compared to manual interaction (He, et al., 2015).

The main factor of increased cognitive demand is thus believed to be interaction time. High levels of mental workload has been observed when several steps are required to complete a task and when system errors occur, prolonging duration of the task. Well-functioning voice systems should help drivers keep their eyes on the road without demanding significant cognitive efforts. A poorly designed system can impose a direct safety threat as it could force the driver to switch focus from the road. A voice system designed for in-car usage must demand less operational effort than its manual counterpart (Cooper, Ingebretsen and Strayer, 2014).

Response time is another factor that has been proven to affect the cognitive demands of the user – which is especially important in a driver’s situation. Responses that are delivered too quickly put more pressure on, confuse and frustrate the users as it becomes difficult to process information (Klein, 2015). Responses with too long delay times should also be avoided as it may cause the user to feel stressed while waiting for feedback and allocate too much attention on the system rather than driving. Delay time below four seconds has been suggested to be optimal for systems to be used while driving (McWilliams, et al., 2015).

To sum up – voice is suitable as a modality since speech is a natural means of communication, giving a system input by voice can be faster than typing and it allows the driver to keep their eyes on the road. However, voice interaction in a car setting does pose some cognitive distraction on the driver and thus it is important that the voice interaction is kept short, clear and contains as few errors as possible. A poorly designed voice user interface poses too much cognitive load on the driver, which prevents safe interaction while driving.

### 3. METHOD

The research method used to conduct the thesis project is presented in the following section. It begins with the chosen design approach followed by purpose and execution of a questionnaire and design workshop. The questionnaire explored current behavior for controlling music applications while driving while the design workshop focused on designing dialogues between user and music service and understanding expectations of such a voice user interface.

An exploratory design approach was used as it is a suitable way to tackle design tasks in new fields (Martin and Hanington, 2012) and since no particular design process that could be applied to this design task (Lynn University Library, 2015) – other than general guidelines on designing voice user interfaces – was found in the literature research. The focus was to understand users’ challenges, needs, desires, interactions, preferences and use patterns in order to gain a comprehensive understanding of the target users and area (Martin and Hanington, 2012). Findings from complementary methods are in an exploratory approach combined to understand both the user and existing systems before resulting in tangible design implications and guidelines, an approach which was suitable for this design task as it included design challenges from different areas in order to provide a safe and enjoyable user experience.

A questionnaire and a design workshop were conducted to involve users in this design process.

#### 3.1 Questionnaire

For this study, a short questionnaire was conducted with the purpose of understanding what music controls drivers use in their music applications and what controls they avoid as well as why. The responses of the questionnaire was the starting point of the design task as it helped understand the user’s part of the dialogue. It was considered important to understand what drivers think about interacting with their music applications before deciding what type of commands the music service should support. The questionnaire can be found in Appendix A.

The respondents were asked to rate on a 5-level Likert scale how likely they were to use ten different commands in their music applications while driving and then in a list check those controls that they choose not to use, along with commenting on why they decide not to use them.

To find potential flaws in the questionnaire before distributing it, a few university students and a user research employee at Spotify were asked to read it and provide feedback on questionnaire format, questions and response alternatives. The questionnaire was distributed on Reddit’s SampleSize community, several Facebook groups for car enthusiasts and a Swedish online car forum. Respondents of interest were people with a driver’s license who listen to music applications – Google Play Music, Spotify, iTunes or other smartphone music applications – in their cars. Age was not considered an eliminating factor as the aim was that the results of this study would be useful for everyone with a driver’s license. However, to make sure that the responses received were useful to the study two eliminating questions were included, asking respondents if they had a driver’s license and how often they listened to music applications while driving. Since people without a driver’s license have most likely only been in a passenger situation in the car, they would not know what it is like to control music applications while driving and thus their responses was not considered to be relevant for this thesis study.

#### 3.2 Design workshop

In order to design the music service’s part of the conversation in a way that provides users with a feeling of familiarity and gain an understanding of how drivers would like to interact with a music...
service – more specifically Spotify – through voice, a design workshop was conducted with Spotify users. A design workshop was considered suitable for given that it offers an efficient and fun way to gain creative input and valuable insight from participants (Martin and Hanington, 2012).

Five users participated in the workshop, which was held at the Spotify HQ. The participants were recruited through social media, online forums and through Spotify employees. The requirements to participate in the workshop were for the users to frequently listen to Spotify and have a driver’s license. The participants should also have listened to Spotify several times while driving and should have tried out a voice-controlled application at least once. Since a goal of the workshop was to investigate the familiarity aspect and get concrete examples of what a dialogue between driver and Spotify could sound like, which is why it was important to have participants familiar with Spotify and its brand. As with the questionnaire, it was important for the participants to know what it is like to control music applications while driving. The participants were between age 25 and 42, three were female and two were male. The participants worked within IT, digital media and home care service.

To get the participants started with the design workshop and make them feel comfortable in the setting, they were asked to individually write down thoughts on voice interaction onto post-it notes. After five minutes of writing, each participant were asked to put their post-it notes on a whiteboard and share their thoughts with the rest of the group. This activity was followed by a short brainstorming session on “What words come to mind when you think about Spotify?” and “What do you think about listening to Spotify while driving?”. The participants’ thoughts were written on the whiteboard next to the post-it notes during the brainstorming session. The purpose of these two activities was to have the participants start thinking about voice interaction, the music service Spotify as well as listening to music applications while driving before moving onto the actual design phase.

During the design part, the five participants were split into two groups of two and three – Group A with participants A1 and A2, and Group B with participants B1, B2 and B3. Each group received one story for which they would design two different dialogues between driver and Spotify. The stories were chosen based on findings in the literature presented in section 2, the questionnaire and elements from the music client. The three stories, presented by their titles, were

1. Do you have any good travel playlists?
2. Play me something like this.
3. Oops. There was an error.

Story 1 was about finding a playlist suitable for travel. In Story 2, the driver listens to an album and wants to listen to something similar after the last song of the album is done playing. In Story 3, the driver mispronounces the name of a song. The stories covered search, error handling, presenting options and search results, and pivoting a context. The story descriptions given to the workshop participants can be found in Appendix B.

The participants were encouraged to take inspiration from the first part of the workshop while designing their conversations, and were also provided with screenshots of the Spotify Android application. Group A worked on Story 1 and Group B on Story 2. After 30 minutes of designing, the groups were asked to present their conversations to the other group and explain how they had been reasoning around their story. After the presentation, the activity was repeated. This time Group B received a new story, Story 3, while Group A was assigned Story 2 for a second iteration. The groups spent 20 minutes on the second design task and they were encouraged to take inspiration from the previous discussion.

The initial plan was to split the participants in three groups so that each story would be iterated on a second time, but due to last-minute dropouts caused by illness the plan had to be altered. It was decided that Story 2 would be kept for a second iteration as both Story 1 and 3 included searching for something specific whereas Story 2 was the only one starting from a context.

After presenting the newly designed dialogues, the participants were given five small post-it notes each and were asked to place the post-it notes on the different conversation components and actions – prompts, grammars and dialogue logic – that they liked the most. The voting helped give a better picture of what a dialogue between driver and music service would preferably sound like and what parts the design task should focus on.

The thesis student facilitated the design workshop – giving instructions and leading discussions forward if needed – and the entire session, which lasted for 2 hours, was recorded.

### 3.3 Designing voice interactions

After the questionnaire and design workshop, a set of voice interaction suggestions was designed with decisions backed up by guidelines and findings in previous research, questionnaire responses and design workshop. Design guidelines at Spotify were taken into account during the design process.

### 4. RESULTS

The results section presents findings from the questionnaire followed by results from the design workshop. The design workshop included a brainstorming phase and two design phases where the participants designed dialogues for three different stories. A vote on conversation components and actions ended the design workshop.

#### 4.1 Questionnaire

88 responses were received from respondents located in Australia, Canada, the Netherlands, the UK, New Zealand, Israel, Sweden, and USA. 9 responses were discarded as one respondent did not have a driver’s license and the other 8 stated that they never listened to music applications while driving, making it a total of 79 valid responses. A majority (69) of the respondents stated to be driving daily and more than half of the respondents listened to music applications always (33) or often (29) during their ride.

When it comes to avoiding the use of controls – see Appendix A for the controls – the five search types were the most avoided followed by switching to a new listening context, as presented by Figure 1. Many respondents mentioned that they never base their search on year, contributing to the high avoidance number for that search type. For the other four search types the main reason to avoid them was safety concerns; searching takes focus from the road and it takes a long time to type search queries. One
respondent who avoided switching context and using all search options except for playlists explained that

“I don’t use them so often when I drive because I’m a slow typer when I use my phone. It takes me too long to type an artist’s/band’s name and search for the specific album I want to listen to. It takes too much distraction away from the road.”

Several respondents mentioned skipping through songs in shuffle mode to find what to listen to, an activity which can be cumbersome when looking for a specific song or wanting to match music with current mood. One respondent explained on the subject that

“I prefer not to faff around or be seen by other drivers looking down at a phone whilst driving, unless I am sat in stationary traffic. So I just tend to press "skip" in shuffle settings to find better songs, which can take forever depending on my mood. I search via song only as my phone will still play other random songs afterwards, whereas if I selected an artist/album it’ll only play that music. So if there’s only one song in an album I like, I have to listen to unfavorable music until I stop to change it.”

In short, three main reasons why the respondents chose not to use certain music controls could be found – safety concerns, some controls can be used with existing buttons in the car, and thirdly that the controls never were being used by the respondent at all.

Another takeaway from the questionnaire was that both voice interaction and a car mode for music applications were something that drivers were wishing for. The questionnaire participants stated a wish for the ability to easily say what song to listen to without relying on a passenger to control the application. Having an easy-to-use interface to interact with was also desired as existing interfaces were perceived to be too tricky to use while driving.

4.2 Design workshop

4.2.1 Brainstorming phase

The results from the design workshop are presented below, starting with the brainstorming phase on voice interaction, the music service and listening to the music service while driving.

When presenting their post-it notes after having the participants thinking about voice interaction, positive aspects that were brought up were that voice recognition is good when it works and that it is the perfect choice in hands-busy situations. On the contrary, the participants felt that interpretation often is bad due to for example mispronunciation and that it can take too much time to complete a task. A requirement from the participants about voice interaction was that it should be easy to go back a step or start over if needed without always having to redo all steps required to complete a task. The participants wished for the voice output to be like a friend, without being overly polite, and that the voice should resemble a human’s rather than being computer-like. Another wish was to be able to view progress on a screen during interaction.

When it comes to the music service Spotify, the participants viewed it as both an emotional and a musical diary. One word that was mentioned to describe it was happiness and one participant stated that “It is a part of my everyday life!” Spotify provided the participants with freedom and easy access to new music; it was considered to be a gold mine in terms of music.

The experience of listening to Spotify while driving was described as “the best thing in the world”. However, for short trips it was considered easier to listen to FM Radio than to set up an AUX or Bluetooth connection with a smartphone. When it comes to content, it was considered easier to find FM Radio channels based
on genre than finding a genre in the Spotify application while driving. “It is easier to access P2 [Public service radio station in Sweden] than your playlists or a genre in the app”. A thought that was brought up next was that maybe Spotify should suggest music for the driver if requested.

4.2.2 Design phase
Below are the dialogues that the participants designed during the workshop for the three different stories. The participants shared their thinking behind each dialogue when presenting to the group, and those thoughts will be presented below followed by the dialogue which they concern.

4.2.2.1 Story 1 – Do you have any good Travel playlists?
Group A based their designs on the wish to be the one in charge of the conversation, especially when driving. “We don’t want Spotify to nag us all the time with questions like ‘So, do you want to listen to some music?’”, participant A1 explained. Participant A2 filled in with “In the car, we want to be the ones saying ‘Hey, I want to listen to something’.” The first dialogue for Story 1 tackled the situation where the driver has a playlist with the same name as a travel playlist, and in that situation the driver should be able to choose if it is the own playlist or a pre-made one by Spotify that should be played.

Story 1, dialogue 1
Driver: Play travel playlist.
Spotify: Do you want to play from travel category or from own playlist?
Driver: Own | Travel category

In the second dialogue prompts for the driver’s mood and should choose a suitable playlist based on that. However, replying with your mood raised some questions by the participants. “What if I’m sad – do I want to listen to something sad or something up-tempo?”

Story 1, dialogue 2
Driver: Play travel playlist.
Spotify: What’s your mood?
Driver: I’m sad, play me something sad | I’m sad, play me something fun
Spotify: Now playing ‘Nu rullar vi’ from travel category.

Figure 3 – Group A presenting Story 1 after the first design phase.

Another thing that was considered to be of importance was the possibility to easily make changes, as shown in the dialogue below. “If I want to listen to some other playlist, I should just be able to say so.”, participant A2 said. When changing music, Spotify should not remember the mood that it was given previously. Participant A2 explained since one’s mood might change over time. Participant A1 filled in with “That’s tricky. I don’t want it to ask me all the time ‘Are you happy now?’”. The music service should be able to make some choices on its own and participant A1 explained it as “It would relieve you from decision agony and you won’t end up in a circle of ‘I don’t know what I want to listen to’”. Participant A2 filled in with “It really should just find something else to play and not always ask me what I want to listen to or always present me with options.”

Story 1, dialogue 3
Driver: Change playlists.
Spotify: Changing to Family Road Trip.
Driver: Give me other options.

4.2.2.2 Story 2 – Play me something like this.
Group B also worked from the idea of them being the initiator of the dialogue. “It makes me mad if someone interrupts me when I’m concentrating on driving.” participant B1 said. In Group B’s first dialogue, the driver can either say “Play playlist, album or artist related to Muse” or “I like this one, can you play me something similar?” and Spotify will respond by playing short previews from a few songs for the driver to choose from, participant B1 explained. Spotify should then prompt for what type of context the song should be played in – such as an album or a playlist. Participant B3 explained that “You don’t want to end up at not knowing what to listen to again. You don’t want to hear just the song – but rather the entire playlist, album or artist’s collection.”

Story 2, dialogue 1
Driver: Play playlist/album related.
Spotify: I liked this one. Play something in similar mood.
Spotify plays 5 seconds previews of songs
Driver: I like “this one”.
Spotify: Do you want to listen to the album, artist or this genre?

The second dialogue is based on the driver’s mood. Participant B2 explains that if the driver for example feels like “I’m really in a party mood – play me something party!” the music service could respond “These are my favorite party songs” maybe based on the most popular party songs at the moment. As the dialogue is mood-based, participant B3 explained that the driver should be able to steer the music suggestions based on mood, feelings or genre by saying things like “Faster”, “Slower”, “More house”, “I’m sad today – sadder!”

Story 2, dialogue 2
Driver: I’m in a party mood. Keep this up!
Spotify: These are my favorite songs.
Driver: Save the playlist | Maybe something more like house | Faster | Slower

Another requested feature was the possibility to save songs in a playlist. “If I like the songs that are playing, if I for example think they go well together, I would like to save them in a playlist.” participant B2 explained. Participant B3 filled in with “It’s difficult to remember the songs afterwards and I don’t want to
think about it when driving, nor do I want to go to my app to save the songs”.

Figure 4 – Group B working on one of their stories. Whiteboard from brainstorming phase in the background.

Group A, which did the second iteration of Story 2 worked with the idea that the music service should start playing the next album by that artist. A1 explained their reasoning behind that choice as “That artist is what you chose to listen to – so why not just continue?” “Spotify will tell you ‘Now Playing ___’, and from here we will be in charge again.” participant A2 continued. Participant A1 filled in with saying that if the driver would like to listen to something else the driver should be able to say “Play related artist or album”.

Story 2, dialogue 3
Last song of an album is done playing. Next album by that artist starts playing automatically.
Spotify: Now playing ___ by Muse.
Driver: Play related artist.

The other way which Group A worked with to pivot the context was to use the radio station feature, which provides the user with similar music based on an artist, album or song. “If you want to get even further away from what you were listening to from the start, you should be able to start a new radio station from the new songs that Spotify plays.” participant A2 said followed by “You should for example be able to say ‘Play only this artist’”. “The radio feature can easily provide the driver with new music. It’s a little bit like exploring related artists except that it doesn’t choose directly from a playlist. This time the results will depend on from where you make the choices, it will be like listening to a nice mix tape. This doesn’t put that much effort on the driver as the radio feature will fix most for you if you want to hear something similar.” participant A1 explained.

Story 2, dialogue 4
Spotify: Now playing ___ by Muse.
Driver: Play Muse Radio.
Driver: Play only from this artist.

4.2.2.3 Story 3 – Oops. There was an error
Participant A1 presented their thinking behind Story 3. “If you search and Spotify doesn’t find the song you are looking for but picks up the artist, then it should suggest ‘Did you mean Uprising by Muse?’ If you respond ‘Yes’, Spotify should ask if you want to play or queue the song.” If the suggested song was not the one the driver intended to search for, the driver should be able to say “No, I meant...” followed by the correct name.

Story 3, dialogue 1
Driver: Search for ‘Upspring’ by Muse
Spotify: Did you mean Uprising by Muse?
Driver: Yes | No, I meant...
Spotify: Do you want to play or queue this song?

In case Spotify could not pick up neither artist nor song, it should let the driver know and ask for the driver to repeat the search query. “And then we came up with the idea that you should be able to interrupt Spotify in case you’ve learnt what it will say – you shouldn’t have to listen to a novel.”

Story 3, dialogue 2
Spotify: I couldn’t hear what you said. Please repeat the artist.
Driver can use a shortcut to get through the dialogue.

Group B also expressed their wish for a quick mode where the driver gets auditory feedback instead of words. “It could be difficult to hear full sentences properly in the car, the sounds might be easier to distinguish. And you might want a shorter path through the dialogue if you are a younger driver than if you are like 72 years old.” participant B3 said.

4.2.3 Voting – most liked conversation components and actions
The 25 votes were distributed as follows

- 5 votes
  - Save playlist
- 3 votes
  - Automatically playing the next album by an artist once the last song of an album is done playing
  - Quick mode with auditory feedback instead of voice output
- 2 votes
  - “Now playing ___”
  - “Do you want to play or queue this song?”
  - “Did you mean Uprising by Muse?”
  - “Maybe something more like house”
  - “Play Muse Radio”
- 1 vote
  - “What’s your mood?”
  - “I’m sad, play me something sad.”
  - “Do you want to play from travel category or your own playlist?”
  - Ability for the driver to interrupt Spotify when speaking

4.2.4 Additional workshop findings
At the end of the workshop the participants were asked if they would like to use a voice-controlled Spotify based on the conversation components and actions that received the most votes, and all participants said that they would use that voice user interface.

Other thoughts that were raised by the participants were the use of a wake word – “Do I have to say ‘Hi’ to the app every time?” – that it should be easy to fall back on typing if needed and that it
would be desirable if the driver was able to choose voice for the music service. The participants mentioned that they want to be able to choose if the voice should be male or female, as well as its dialect. “Like, a drunk Irish guy when you are in a party mood. That would be fun! Especially when you are in a talkative mood”, said participant A1 who also wanted the voice to be able to follow the driver’s conversational style. “If I don’t want to chat, it should understand that and just listen.”

5. VOICE INTERACTION DESIGN SUGGESTIONS

This section presents voice interaction design suggestions based on previous literature, findings in the questionnaire and design workshop. The design suggestions include both absolute semantic and context-dependent commands, which were introduced in section 2.2.1.

5.1 Absolute semantic commands

When interacting with a voice user interface, users should always be able to fall back on absolute semantic commands (Dahlgren, 2013). While some questionnaire participants stated that they use hardware buttons in the car for some controls, one participant did express their interest in being able to tell the music service to stop playing instead of looking for a button. With that said, simple commands like “Skip song” and “Mute” should be supported for drivers who prefer uttering them. Paraphrasing of commands needs to be taken into consideration (Schalkwyk, et al., 2010), meaning that a driver should be able to stop playback by saying for example “Stop playing”, “Pause” or “Pause music”. Supporting paraphrasing is especially important in situations that require quick actions.

5.1.2 Context-dependent commands

An overall aim when designing the following suggestions was to support context-dependent commands and the use of natural language. The voice output has been designed to speak in a friendly, short and clear way based on the wishes of the design workshop participants for the voice output to be friendly. Guidelines on in-car technology and voice interaction design have been taken into account.

5.1.2.1 Decreasing interaction time, ensuring guiding prompts and presenting options

To decrease interaction time between driver and music service for a lower cognitive demand on the driver (Cooper, Ingebretsen and Strayer, 2014) some conversation components and actions were removed from the dialogues designed during the workshop. For example, the step where Spotify asked for the driver’s mood in Story 1, dialogue 2 was removed, as can be seen in Figure 5. Would the music service always prompt for the driver’s mood, it would not only require the driver’s attention for a longer time but it would also require the driver to always think about what type of music they want to listen to in terms of mood or genre. As the workshop participants mentioned that they wanted for Spotify to make smart choices when choosing what to play, the playlist choice could rely on factors like weekday or time of day.

In the scenario where the search query can result in more than one option, as brought up by the workshop participants, the music service should ask about what to play. This does not only apply for playlists as in Figure 5, but also in situations where for example two different artists have a song with the same name. When asking the driver what option was the intended one, only two different options will be presented to decrease the cognitive load on the driver. The Alexa Skills Kit Voice Design Best Practices, which was presented in section 2.2.1, suggests to not present more than three options to a user. As the user in this scenario is driving, presenting only two options as in Figure 6 will put less effort on the driver to focus on listening, remembering options and making a decision. The same reasoning backed the decision to remove the action in Story 2, dialogue 1 where the music service would play previews for the driver to choose from. Questions have in general been re-designed to prompt the driver for an answer in a way that the driver will know what type of answer is requested – either “Yes” or “No”, an option or additional information.

Figure 5 – the driver is prompted to decide what playlist to play, one of their own playlists or one from the travel category.

Figure 6 – the driver is being presented with two new playlist options to choose from.

5.1.2.2 Music based on mood and genre

While the mood prompt was removed from Story 1, dialogue 2 it was not discarded completely as it was one of the prompts that received votes during the workshop. In the scenario where the driver does not know what to listen to at all and just asks the music service to play something, the music service could prompt for a mood or genre as can be seen in Figure 7. The workshop participants did suggest that the music service should recommend music for the driver, and with a response including preferred mood or genre the music service can provide more suitable music than if it just picked something by random. Since the workshop participants raised the question of “What if I’m sad – do I want to listen to something sad or something up-tempo?” the mood/genre
prompt has been designed in a way which prompts the driver to think about what mood or genre they would like to listen to. Most importantly, it might not be clear to the driver that a response like “I am sad, play me something sad” is requested if the music service only asks “What’s your mood?”.

The driver could also include a mood or genre in their search query and then directly get a new playing context without being presented with different options, as requested by the workshop participants and presented in Figure 8. In the same manner, in case a driver simply wants to search for a song – as requested by a questionnaire respondent – simply telling the music service to play a specific song would be the fastest way to achieve that.

Figure 7 – the music service should ask questions when in need of more information to deliver suitable music. An example is when the driver does not know what to play.

Figure 8 – a simple mood-based search.

5.1.2.3 Queue and error handling
The music service should only ask the driver if the song should be queued or played – as was part of Story 3, dialogue 1 – if the driver asks the music service to search for content as in Figure 9. If the driver on the other hand says “Queue“ or “Play” those functions should be triggered. The reasoning behind this decision was that the driver should not have to answer a sequence of questions every time they want to listen to something, especially since adding additional questions to every search query will increase interaction time and potentially frustration. In case of error, the voice output should also prevent the driver from getting stuck in a loop of errors.

5.1.2.4 Context awareness
Regarding the second iteration of Story 2, where the driver can for example ask for a related artist or start a radio station, nothing has been changed. The driver should be able to both start a radio station based on the current playing context and a specific song, artist or album as presented in Figure 10. Another conversation component that was kept from the design workshop was the “Now playing”, as it received votes and as voice design guidelines suggest that the user always should know what the current context is (Amazon, n.d.b). These notifications have different phrasing depending on if a song was queued, if songs are being shuffled, if a radio station has been started or if an album, playlist or song is simply being played.

Figure 9 – an example of error handling and prompt from music service when the driver uses the command “Search”.

Figure 10 – different ways for the driver to get music based on the current playing context.

6. DISCUSSION
In this section the research is discussed from four different perspectives – the design suggestions, the research method, ethics and future work.

6.1 Design suggestions
Car and voice design guidelines and input from research participants were taken into account when designing the voice interaction suggestions. An interesting takeaway from the workshop was that many of the participants’ wishes were aligned with the design guidelines for in-car technology and voice user interfaces. For example, participants wanted to be in charge of interaction and keep interaction short. It was important for the workshop participants that the music service would not interrupt them while driving. The participants did not want Spotify to ask
too many questions – questions should only be asked by the music service if required, as also suggested by the Alexa Skills Kit Voice Design Best Practices, and in that case questions that will improve the music suggestions should be asked.

6.1.1 Safety concerns
While some of the dialogues were fairly straightforward to finalize, like Figure 10 in section 5.1.2.4, there are some situations that could require some additional attention before implementing with confidence and thus are worth to discuss further due to safety concerns.

6.1.1.1 Error handling
One of the more difficult situations to design properly is when an error occurs. It is not possible to rely on every input to be recognized fully by the system, as the recognition rate at the time of writing was about 90% (Klein, 2015). Thus, errors will most certainly occur during interaction. It is important that the error message provided prevents the driver from getting stuck in a loop of errors and that the error message assists the driver to find desirable music – both from a safety and user experience point of view. In case the driver does not know the name of a song and tries to search for it, but just gets stuck in a loop trying the same song over and over again, frustration will most likely happen. The driver should not have to fall back on interacting with a potentially tricky graphical user interface nor be left frustrated due to lack of music as music plays a large role in creating a pleasant auditory environment.

6.1.1.2 Keeping interaction time short and minimizing user frustration
Keeping interaction time short is a very important aspect to lower the cognitive demand of the driver (Cooper, Ingebretsen and Strayer, 2014) and an implementation using voice control should require less effort than its manual counterpart (Cooper, Ingebretsen and Strayer, 2014; ustwo, 2014). One situation where this might not be the case is the one presented by Figure 9, section 5.1.2.3. Except for the error handling prolonging interaction time, as discussed in the previous sub-section, how to handle drivers using different phrasings such as “Queue”, “Play” and “Search” in terms of interaction time is also a concern. Especially when an error occurs.

The suggestion was made in section 5 that the music service only should ask if a song should be queued or played when the phrase “Search” is being used. Such a prompt adds interaction time, especially in the situation where there is an error. Having the music service doing one or the other by default without asking might leave the driver frustrated if for example the current playing context is being interrupted with a new song when the driver actually wanted to queue the song. Such a situation might cause the driver to pick up their phone or initiate a new dialogue with the music service – two scenarios that steal focus from driving.

As presented in section 5, the mood prompt was removed from Story 1, dialogue 2 to keep interaction time short. It is better to teach drivers that they can do searches on mood or genres directly rather than having the music service prompting for it every time. This requires a thorough training session and some effort from the driver before stepping into the car, but will decrease interaction time and the cognitive load on the driver. Matching music with mood has been shown to be very important as presented in the introduction and brought up both in the questionnaire and the design workshop.

6.1.1.3 Presenting options
Presenting options, as in Figure 6, is also a tricky topic. As the song previews from Story 2, dialogue 1 was not part of the design suggestions in section 5, it could be worth to evaluate which option – previews or titles – would pose the least cognitive demand and requires the least attention from the driver. While the driver is being presented with different playlist titles that to some extent might explain the content of the playlists, they are essentially just playlist titles and do not necessarily say much about the content of the playlist. It could be worth evaluating if the driver should hear a few previews from the playlist or if it is better for the driver to switch to another playlist in case the driver is not satisfied, as proposed in the workshop in Story 1, dialogue 3. While the first suggestion would require more attention from the driver, the second one might cause more frustration if the driver needs to ask for another playlist several times.

In the scenario where two different artists have a song with the same title and the driver does not know the name of the artist, then playing a sound bite of each song could be a better way of helping the driver decide what song to play. That could be a more suitable situation for previews rather than for a playlist, which is a collection of songs. Only playing the first song of the playlist in order to keep interaction time shorter might not present the driver with a correct representation of the entire playlist.

6.1.2 Familiarity with the music service
One important aspect of the thesis study was for the voice interactions to provide the users with a feeling of familiarity with the music service and its brand. It has been suggested that giving the voice output a personality and aligning it to the brand would provide a better user experience (Klein, 2015). The workshop participants stated that they viewed the music service as an emotional diary, that it was associated with a feeling of happiness and part of everyday life. Music is thus not only very important to people and does not only act as an auditory decor – the service providing the music plays a large role as well.

Several of the dialogues designed during the workshop included features available to Spotify users, such as the radio feature. The workshop participants also wanted to be able to perform actions related to their account, such as save songs in a playlist. Being able to perform the same actions through voice as they normally would, such as adding songs to their personal music diary, could be a main factor that differentiates this music experience from for example listening to FM Radio – thus being the familiarity factor.

The phrasing of the voice output can also contribute to the feeling of familiarity. The workshop participants mentioned that they wished for the voice to be like a friend, and one of the conversation components was the music service suggesting “its favorite songs”. One aspect that was brought up several times during the workshop was that the music service should know the user. Knowing the user well is out of scope in terms of voice interaction design as it depends on underlying programming factors, but as it was considered so important it is a thing to keep in mind for future work.

An interesting contradiction was that the workshop participants requested the ability to interrupt the music service, as Klein
(2015) suggested that interrupting the speaking output would leave the user feeling impolite. It could be useful to have such a command in a car setting though given that the participants said that they did not want the music service to interrupt them and guidelines by the European Commission (1998) suggest that the driver should be in charge of the interaction. The ability to interrupt the music service could also be helpful in situations when the driver wants none of the options provided by the music service and would like to cancel the current interaction completely or start over. If the interrupting command would be for example “Quiet” it would also require the music service to understand if it should stop talking or stop playback — a typical situation where the current context matters. However, it would require the music service to be able to take input while it is busy with output and not end up in a loop of errors as suggested by Klein (2015).

Even though the design workshop provided initial insight into what the feeling of familiarity with Spotify could be, it is an aspect that could be explored further. As only five workshop participants, who were frequent Spotify users, input could be gathered from a broader group of Spotify users with various usage frequency to gain a better understanding on how brand recognition can be achieved through voice interaction.

6.2 Research method
The questionnaire and design workshop complemented each other well when approaching the research question. The questionnaire provided insight in current driver behavior in terms of using music applications while driving. The design workshop resulted in concrete examples of what a dialogue between driver and music service could look like along with additional insights on what users think of voice interaction and how they perceived the music service. Together, the questionnaire and the design workshop provided a good understanding of what drivers expect from a voice user interface for their music application.

6.2.1 Questionnaire
The results from the Likert scale on likeliness and those from avoidance of music controls did not correlate fully, but still gave an indication on drivers’ usage of music applications while driving. Some respondents also contradicted themselves, saying for example that they were likely to use a search type but then stated that they avoided the same search type. Those respondents were not disregarded in the results though as they still provided valuable qualitative data in their written response.

The different nature of the questions in the questionnaire can be an affecting factor on the slight difference between the numbers in likeliness and avoiding use of controls. Firstly, the responses to the Likert scale were merely approximations given that they were asked in a questionnaire and not in the actual scenario that they regarded. Secondly, respondents could possibly have stronger feelings about avoiding controls in a scenario that requires a lot of concentration than when being asked to rate likeliness on a Likert scale. Having both types of questions and combining them with written responses provided a good overall picture of controlling music applications in the car. With that said, the questionnaire provided adequate data to understand driver behavior and a first insight in the driver’s part of the conversation.

Search by genre was not included in the questionnaire but as it was brought up in the design workshop it was considered important and was thus not excluded from the thesis study.

6.2.2 Workshop
Due to last minute dropouts prior to the design workshop, fewer participants than planned attended. Since the initial plan had to be altered slightly, only one of the three stories was iterated on a second time. While it would have been interesting to see the results of a second iteration on all three stories, the design workshop did provide interesting results to work with for the design suggestions. Most importantly, the participants were satisfied with the output and stated that they would use the voice user interface that they had designed dialogues and voted for.

6.3 Ethics
6.3.1 In-car technology
As the use of technology while driving is not completely risk-free, a topic worth discussing is if it is right to develop technology to be used in a situation that requires significant attention from the driver. Since drivers will continue to use technology to some extent it is better to support the driver accomplish their task with easier-to-use interfaces than having them interact with interfaces that are not suitable for driving. Both previous research and this thesis study have shown that drivers are very concerned with safety and that voice-based applications and easy interfaces are something that they long for.

It is important to keep in mind that when designing in-car technology a huge responsibility is put on designers and developers to ensure that guidelines are being followed. Evaluation is of great importance to find improvement points before a release as a poorly designed in-car technology is a safety threat. The technology should assist the driver, not cause accidents.

6.3.2 User participation
The anonymity of participants was respected throughout the thesis work. Questionnaire respondents were informed what kind of research their responses would be part of and no personal information about the participants was collected except location and age. To minimize the risk of workshop participants feeling uncomfortable, they were informed already when signing up for the workshop that it would be recorded and that photographs would be published in the thesis report. The workshop participants were also assured that their names were not to be published and that no one else than the author of this report would have access to the recording.

6.4 Future work
The next steps in a project like this would be to implement a prototype, evaluate and improve the voice interactions. Questions to base the evaluation on could be: Are drivers provided with adequate music results?, Can tasks be accomplished in few steps with minimal errors? and Is the dialogue effortless, i.e. is it easier to control the music service by voice than its manual counterpart? Evaluation should first be done in a car-like setting — perhaps in front of a screen or in a simulator — so that all pain points have been covered through iterative design before evaluating in an actual car setting to prevent stressful situations and accidents during evaluation.

Additional possible extensions for future work concern localization, which was outside the thesis scope, and the voice to use. Things to investigate further are how phrases should be
translated to still work well in different languages and if the music service should have only one voice for output or if the driver should be able to choose. For the choice of voice, it would be interesting to see what provides most familiarity for the driver. Maybe a drunken Irish voice really increases the user’s likeability compared to a generic voice. Since people apply biases depending on the gender of the voice (Jeong and Shin, 2015), it could be better to let the user decide voice based on their preferences or idea of what gender Spotify is to them.

The wake word used to initiate interaction should be evaluated in a car setting to see how it works in a setting with several people speaking or with other systems also listening for wake words as systems should not pick up contexts that are not meant for them (Dahlgren, 2013). The wake word should thus be designed in a way that other systems, like the GPS, cannot be triggered and so that everyday conversations do not trigger the music service as it might interrupt the driver. An easily accessible button could be placed in the graphical interface to support multimodality, especially since the workshop participants expressed their concern about frequently needing to use the wake word.

The workshop participants stated that they would like to be able to view progress on a screen, easily be able to fall back on typing and suggested that it might be easier to listen to sounds than words when driving. Visual and auditory feedback and how to design them are two fields that were outside of the thesis scope but require further attention – especially how to balance them in order to minimize cognitive demand from listening or reading. Nielsen (2003) suggests that voice works best in a multimodal setting, but it might not apply when the user is driving as visual distraction has a large impact on driving performance (Peissner, Doebler and Metze, 2011). While it is difficult to listen to a lot of information due to its sequential and non-persistent nature (Schalkwyk, et al., 2010), it requires less cognitive effort to listen than to read while driving. Visual feedback should thus only require short glances from the driver. It is important to find a proper balance between visual, auditory and spoken feedback in order to provide a safe and pleasant user experience that does not steal too much focus from the driver.

The conclusion section summarizes the findings of the thesis project. The findings are presented through the two sub-questions and the research question that guided the research, which were introduced in section 1.2.

What controls do users want to use when driving?
Drivers want to easily and safely find the specific music they want to listen to or music that suits their current mood without needing to rely on a passenger. An interest exists in talking to a music application and existing features from the music service should be supported as voice commands. Drivers want to be able to quickly tell the music service to for example pause the music. The voice-controlled application should also be combined with Bluetooth integration for controls on the steering wheel and the possibility to interrupt the speaking voice was requested.

How should the music service communicate with its users?
The driver should be in charge of the conversation – the music service should not break the driver’s concentration. The voice output of the music service should be like a friend – friendly and brief without being overly polite. In case the driver’s input will help the music service present better music options, the music service should prompt the driver for additional information such as what kind of mood the music should resemble. The driver should know what the new playing context is and the music service should give adequate support in case of an error to help the driver find likeable music.

What voice interactions should be designed in order to provide hands-free control of an already existing music service, making sure that the dialogues provide users with a feeling of familiarity with the music service?
Both absolute semantic commands for different phrasings and context-dependent commands should be supported. Drivers should be able to say simple things like “Mute”, using different synonyms. Advanced searches based on aspects like mood, genre or the current playing context should be supported. When it comes to the feeling of familiarity, drivers should be able to use existing functions supported by the music service and complete actions affecting their own accounts like saving songs to a playlist. The voice output should be like a friend who suggests music and knows the driver and their taste in music.

8. ACKNOWLEDGMENTS
I would like to thank Lawrence Kennedy who with great enthusiasm welcomed me to his team to conduct my Master’s thesis.

9. REFERENCES


Amazon, n.d.c Amazon Echo. [online] Available at: <http://www.amazon.com/Amazon-SK705DI-Echo/dp/B00X4WHP5E> [Accessed 1 October 2015].


Appendix A – Questionnaire

Age
- 20
- 21 – 30
- 31 – 40
- 41 – 50
- 51 +

Location
*Please specify as City, Country.*

Do you have a driver’s license?
- Yes
- No

How often do you drive?
- Daily
- Weekly
- Monthly
- A few times per year

How long is your usual drive?
*Please specify in minutes or hours.*

When driving, how often do you listen to music through an app?
*For example Google Play Music, iTunes, Spotify or another smartphone music player.*
- Always
- Often
- Sometimes
- Rarely
- Never

Please estimate how likely it is that you would use the following controls in your music app while driving.

<table>
<thead>
<tr>
<th>Control</th>
<th>Very unlikely</th>
<th>Unlikely</th>
<th>Neutral</th>
<th>Likely</th>
<th>Very likely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Play/pause</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Skip track</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Go back to previous track</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Volume up/down</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Search playlist</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Search song</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Search album</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Search artist</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Search based on year</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Change from current playlist/album/song/artist to something completely new</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>
Are there any controls in your music app that you choose not to use when driving? Please check them in the list below.

☐ All of the controls in the list below
☐ Play/pause
☐ Skip track
☐ Go back to previous track
☐ Volume up/down
☐ Search playlist
☐ Search song
☐ Search album
☐ Search artist
☐ Search based on year
☐ Change from current playlist/album/song/artist to something completely new
☐ None of the above

Please describe why you choose not to use the controls that you checked in the list above.

*If you chose the option "None of the above", please type "Not applicable".*

Any other comments on listening to music apps while driving?


Appendix B – Design workshop stories

Story 1 – Do you have any good travel playlists?

Your story
You are driving and you want to listen to a, by Spotify, pre-made playlist from the travel category without having to fiddle around in your Spotify app to find a suitable playlist.

Questions to inspire you
- How do you ask Spotify to find such a playlist?
- How does Spotify present different options to you?
- How do you choose what playlist to play?
- What if you don’t want to listen to any of the playlists that Spotify presents to you?
- What is the purpose of your drive?

Anything else that you can think of that should be part of the dialogue?

Psst… You can find a few examples of travel playlists in your Workshop Kit.

Story 2 – Play me something like this

Your story
You are listening to the album “The Resistance” by Muse. You realize that the last song is playing and so you would like to listen to something similar – the problem is that you cannot look for that in your Spotify app right now since you are driving.

Questions to inspire you
- Do you want to listen to a similar album, artists or a mix of songs?
- Should Spotify be the initiator of the dialog by suggesting options to you once you are done listening to The Resistance?
- How do you want Spotify to tell you what will be playing?
- Should Spotify give you options to choose from?

Anything else that you can think of that should be part of the dialogue?

Psst… You can find screenshots of The Resistance in your Workshop Kit.

Story 3 – Oops. There was an error

Your story
You want to search for the song “Uprising” by Muse but mispronounce it as “Upspring”, meaning that Spotify cannot find the desired song. Since you are focusing on what is going on at the road ahead of you, you don’t want to look for the song by scrolling through the artist page in your Spotify app.

Questions to inspire you
- How do you search for the song?
- How should Spotify let you know that it could not find the requested song?
- What do you say to Spotify when you did not get the song that you asked for?
- Should Spotify give you any options?

Anything else that you can think of that should be part of the dialogue?