# Crafting an Engaging User Experience for an In-Car Voice Assistant
## A Design Thinking Approach

Nehal Sharma
Department of Computer Engineering
San Jose State University
San Jose, California, USA
nehal.sharma@sjsu.edu

## Abstract

**Voice interfaces have become increasingly common and are extensively used in smartphone assistants(like Siri and Cortana), and home devices(like Amazon Echo and Google Home). Considering the rapid advancements in the automobile industry and the popularity of conversational assistants, the presence of an in-car voice assistant can play a pivotal role in shaping consumer behaviour, enhancing brand recognition, and improving driving experience in cars.The driver's eyes and hands are occupied while driving, therefore, digital interactions like navigating directions and receiving calls, can be best accomplished by voice.Recognizing these use cases makes the concept of an in-car voice assistant a new domain which can be explored to deliver richer and safer driving experiences.**

**Distracted driving is a prominent reason of car accidents, which accounts for 25% of all car crash fatalities.Therefore, in this paper, I have reviewed a prototype for an in-car conversational assistant that reduces visual, manual and cognitive distractions of the driver, thereby, enabling a safer driving experience. The agent is capable of responding to digital devices, providing visual display of information, improving the driver's emotional state of mind and providing navigational assistance.Effective UX tools like competitive analysis, storyboarding, journey maps, and user interviews were used to understand user expectations and create the persona of the assistant and its high-fidelity prototype accordingly.**

**Another important goal of this user experience research was to acknowledge the requirements, psychology and motivations of the user through observations and feedback.Therefore, I evaluated the design decisions of the conversational assistant, through intensive user testing and heuristic evaluation results, to assess how well their speech capabilities align with user contexts from real world scenarios.The quantitative metrics obtained from the test results were further used to enhance the agent's design simplicity, spontantanity, and alignment of the conversation with respect to real life situations. Designing an Alexa skill for the in-car voice assistant, conducting user testing in immersive environments and evaluating user's trust for the assistant based on its gender are the future milestones for this research project.**

**Keywords:** Voice User Interface, Conversational Agents, Automotive Industry, Car Chatbot, Safe Driving, Human-Computer Interaction

## 1 Introduction

The recent advancements in machine learning and natural language processing have significantly shaped the way people interact with devices. Consequently, Voice User Interfaces (VUIs) like Amazon Alexa, Apple Siri, Google Assistant, and Microsoft Cortana have gained immense commercial success.However, majority of the masses associate personal digital assistants with home devices and smartphones only, and the application of voice assistants in the automotive sector goes unrecognized. A recent national survey conducted by Voicebot.ai in January 2020 revealed that the use of voice assistants in a car is a fairly large market.

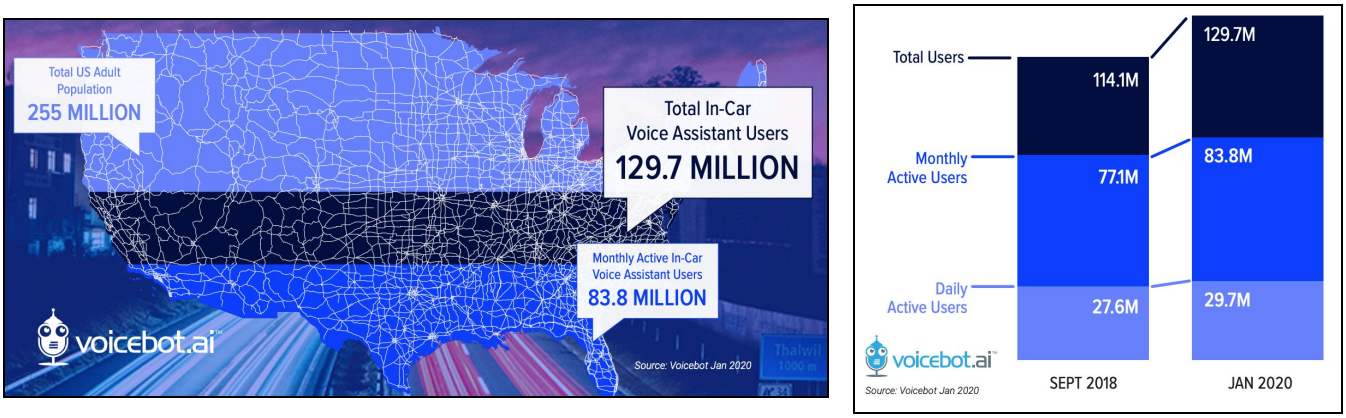


**Figure 1:** In-Car Voice Assistant Total Audience Reach(left) and U.S. In-Car Voice Assistant Users in September 2018 and January 2020(right) (Source: Voicebot.ai, January 2020)

The total audience reach of in-car voice assistants in the United States alone is approximately 130M, which is 51% of the total population.The total users of in-car voice assistants in the United States has grown significantly from 114.1M(September 2018) to 129.7M(January 2020), showing an increase of 13.7%. These figures indicate that more and more consumers are resorting to the use of conversational agents while driving. Thus, the presence of an in-car voice assistant can play a pivotal role in shaping consumer behaviour, enhancing brand recognition, and improving user experience. Not only this, the conversational agent's performance and capabilities can raise the expectations for intuitive interfaces that are simple to use, offer delightful user experience, and deliver accurate results.

This novel idea of conversing with devices may sound exciting, however, using technology to design voice experiences that offer natural and instinctive interactions for everyday tasks is immensely challenging. The lack of human computer interaction theories for designing natural voice user interfaces that meet the user's unconscious expectations and the absence of an architecture for scaling conversations across multiple devices are essential barriers for crafting intuitive speech interfaces.While crafting conversational experiences with automobiles, it is relevant to understand that communication

comes to humans instinctively which makes it extremely complex and difficult to codify. Therefore, incorporating human-like cognitive and emotional capabilities into the conversational agent is important in order to enhance the credibility of interaction between the agent and the user.

Voice Assistants interact via a lot of voice driven commands and audio responses.Thus, they are an offshoot of the ambient computing notion, which emphasises that we are using technology while still engaging with the world around us.In the past, we were accustomed to being behind our terminals, video screens or a mobile device and now, with ambient computing, we are using our voice assistants in our kitchens, while driving or in the living room.The computing devices are still there, but they are not the focus anymore; rather, our lives are the focus and the voice assistants are there to help us live our lives more easily. Therefore, while designing ambient computing systems, like a voice assistant, we need to be more careful about how we are designing our technology and interactions, because people are doing things beyond the machine itself.

## 2 State-of-the-Art Review

Speaking is natural, effortless, and instinctive to humans and people are not going to change their way of conversing with one another, anytime soon.Therefore, conversational interfaces have to be designed to match the user's expectations and to do so, we need to leverage the principles of human conversation.People have been studying and researching language and human computer interaction for a long time and I have utilized this research while prototyping the in-car voice assistant for this project.

Crafting VUIs requires extensive expertise and is not the same as designing graphical user interfaces. Considering the complexities involved in conversation design, it is crucial to understand the recent trends, ongoing research, and the challenges that exist in this field. A sound understanding of these fundamentals will facilitate a more efficient and intuitive user experience of voice assistants. In an extensive study, Stanford researchers, Wei and Landay [1] devised 17 usability heuristics for intelligent VUIs and evaluated three popular devices using these heuristics. This expert evaluation revealed that the heuristics could successfully identify the underlying usability problems and improved the interface design by 18.6%. However, a few research studies fall a bit short in analyzing the human computer interaction paradigms from the design-oriented perspective in the voice domain. Këpuska and Bohouta [2] assess the capability of conversational agents to process combinations of user inputs in different forms such as text, image, speech and sound and also draw a comparison between the knowledge bases of the agent and the user.

It is anticipated that autonomous, self-driving vehicles will radically change day-to-day driving with expected benefits of improved road safety, efficiency, comfort and mobility.The automobile manufacturers are designing next-generation vehicles with cutting-edge technology that allows real-time access of information such as driving directions, nearby locations, emails, alternative travelling routes with less traffic etc to the driver. A virtual assistant, in this scenario, can reduce the cognitive load on the driver's memory to a large extent and provide a seamless driving experience.

While the use cases of a voice assistant in a car is on a rise, many researchers are developing AI powered chatbots specifically for the driving experience. Clark et. al [3] explore whether the presence of conversational agent in a car could engender trust in the technology, based on our understanding of speech in humans.The experiment transported 34 participants in an autonomous vehicle, accompanied by a voice assistant, 'UltraCab' and trust related themes were recognized by studying the conversations that took place between the participants and the AI agent. Other studies by Viswanathan[4] and Comerford [5] study a conversational agent and its capabilities to reduce the cognitive load on the driver's memory by learning the driver's preferences and using it to simplify common tasks that the driver performs.The paper titled 'Why would I talk to you? Investigating user perceptions of conversational agents' by Anne Catherine Saarem[6] proposes a set of guidelines for designing anthropomorphic voice assistants.

For the past two decades, the human-computer interaction community has proposed principles, guidelines, and strategies for creating client interfaces and interaction for applications employing AI inferences. Nevertheless, the diversity of AI designs and eminent cases of failures, show that designers and developers still find it difficult to create intuitive and effective AI powered systems. Hence, "reusable guidelines that can be shared, refined, and debated by the HCI community can prove crucial to validate new VUI designs"[7].Amershi et al.[7] and Doyle et al[8] suggest a filtered set of guidelines from almost 150 machine intelligent related design recommendations, which have been certified through 3 rounds of evaluations on various voice assistants. The development and use of such principles can enhance the design and evaluation of intelligent systems that people can comprehend, believe, and can communicate with effectively.

## 3 Problem Statement

Distracted driving is one of the major reasons of car accidents in the United States, causing around 25% of fatal injuries each year. Multiple factors can attribute to distract the driver while driving.Some of the most common ones include:

- Displacing objects in the car, such as picking up a dropped object on the floor of the car or rummaging through the glove/storage compartment to find something
- Eating food items or drinking beverages while driving
- Talking on the phone while driving using one hand or having hand-free conversation on the phone through a bluetooth enabled device.
- Dialing numbers or texting on the phone or viewing emails or other application notifications
- Distracted by external factors or events such as traffic, animals on the road or pedestrian crossing the street in a wrong manner etc
- Interacting with in-car equipment such as adjusting window panes, locks, front and rear view mirrors and steering wheel etc.

The types of distractions can be classified into three main categories:

- **Cognitive distraction:** Occurs when the driving is physically present but his mind is somewhere else
- **Manual distraction:** Occurs when the drive moves his hands away from the steering wheel consciously or due to recklessness
- **Visual distraction:** Occurs when the focus of the driver's eyes shifts elsewhere from the road

Driving demands constant attention and focus on the road because our brain has to process large amounts of visual information, monitor and predict other driver's movements, and accurately coordinate our hands, eyes, ears and feet to drive safely.Neurobiological research shows that the human brain is not wired to carry out two or more activities simultaneously. Therefore, it processes these activities in a sequential manner by switching amongst the given set of activities that need to be performed at the same time. Attempting to perform two attention demanding tasks at the same time leads to inattentional or perceptual blindness in which the human fails to notice an unexpected stimulus merely due to a lack of attention, rather than any visual defects.

The driver's eyes and hands are occupied while driving, therefore, digital interactions like navigating directions and receiving calls, can be best accomplished by voice.Recognizing these use cases makes the concept of an in-car voice assistant a possible solution which can be expected to deliver a safer driving experience.

# 4 Methodology

## 4.1 Primary Research and User Interviews:

To begin with the design process, the first step was to discover the problem and user pain points. Thus, online surveys and interviews were conducted [15] with people to gain an in-depth insight about people's actions and thought-process while they are behind the wheels. For conducting the online survey, 90 participants aged between 18-70 years were selected and the survey was conducted through SurveyMonkey.com.The questions were categorised into the following three categories:

| Question Category | Insights Gained | Questions asked |
|---|---|---|
| Understand how much time an average human being spends in the car | An average human spends around 6-7 hours in the car daily | • Hours per week spent in car<br>• Main purpose of commuting by car |
| Understand how an average human uses his/her phone while driving | Majority of the participants use phone and keep them in easily accessible locations in the car | • Phone connected or not to the car<br>• Functionalities of phone used while driving<br>• Exact location of your phone |

| | | |
|---|---|---|
| | | while driving |
| Understand the common reasons for getting distracted while driving | Majority of the respondents got distracted due to non driving factors | • The non driving activities performed<br>• The task which distracts drivers the most |

Following the online survey was user interviews , where the participants were categorised in three different age brackets: 17-35years, 35-52years, 52-70 years and each category consisting of 3 participants. The participants represented different professions like students, engineer, doctor, teacher, homemaker, delivery person and retired army officer. Some of the key insights gained from the user interviews were:

- Phones are mostly used for navigational assistance since it is impossible for a human to know and remember all directions.Moreover, Google Maps reduces the load on the memory and is accurate and user friendly.
- Communicating with devices is still in its infancy and people find it intimidating to interact with virtual agents.Some voice assistants are loud, irritating and unable to provide precise and concise information.
- Emotionally disturbed state of mind like stress, fighting with spouse, scolding from boss, traffic etc. just before/during driving affect the driving process significantly.
- The actions that a person has been performing a lot over the past like driving down a known path makes him/her habitual to it and makes him/her more vulnerable to cognitive distraction.

### 4.2 Data Analysis:

After gathering insights from the online surveys and user interviews, I used an affinity diagram to better group the data logically to decipher natural relationships among different contexts.The affinity diagram made sifting through large volumes of interview data easier and encouraged creative thinking patterns. The primary groupings that emerged from affinity mapping were traffic, internal environment, external environment, road rage and careless driving, feeling sleep, drunk driving, emotional disbalance, app notifications, direction navigation, emergency, phone calls, music, and, involuntary actions.

### 4.3 Defining Target Users:

After analysing the user interviews, their opinions, and their age groups, the next step was to create a user person for each of the different age groups and define their specific requirements accordingly. A user persona is a synthetic representation describing how the user interacts with their surroundings and gadgets while in the car.This journey map has been drawn from the user's

perspective to showcase the user goals, expectations, progress, touch points, pain points and opportunities to create impact.

For example, Aleena is an adult female driver, aged 34 years and is an engineer based in Los Angeles.Due to her hectic schedule, she is often tired while driving. Since she has a long commute between home to work and her phone is a distraction, she often finds herself oblivious to her surrounding environment while driving. Therefore, she wants an ideal in-car assistant who can help her stay mentally, visually and physically aware and respond to her phone texts, calls and notifications.

### 4.4 User Needs and Pain Points:

The user personas and journey mapping were useful in identifying the core user requirements and user pain points that were required to push through to obtain a better design solution for improved user driving experience.

The three fundamental capabilities of the proposed design should be as follows:

- Minimize cognitive distraction resulting from factors such as stress, disturbed state of mind and rushing to reach destination on time
- Improve the visual focus by reducing distractions from phone calls, texts, app notifications and direction assistance.
- Provide a richer experience and control with voice commands

## 5 Results

The above mentioned UX research methodologies helped shape the design process so that work could proceed to the next level which is ideation.After the initial interview and survey phase, the scope of the problem narrowed down.So, a secondary research phase was conducted, with the redefined user requirements and pain points.Consequently, the ideation process was started to generate a plethora of design ideas that encompasses a solution to all the user requirements.

### 5.1 Design Solutions through the Ideation Cycle:

**Requirement 1: Minimizing Cognitive Distraction**

**Proposed Solution: Affective Computing**

As per a research conducted by the Affective Computing Group at the MIT Media Lab, computing devices can simulate emotional intelligence, comparable to that of a human, specially when such devices respond to the negative emotions of a human being. Affective computing devices are capable of measuring human emotions by analysing the data collected from the human's facial expressions, voice tone and language, pupil dilation and constriction, and body movements. Designing an affective computing module for our in-car voice assistant could lower down the disturebed negative emotions of the driver by offering calming solutions.

**Requirement 2: Improve the Visual Focus**

**Proposed Solution: Windshield Display of Visual Information**

The driver need to bend his neck if he is viewing navigational instructions on his phone, since, his phone is placed below the eye level; this can cause the focus of the driver to digress from the road, affecting his driving performance.Hence, to overcome this solution, the proposed solution for our assistant is to project the phone screen on a small portion of the windshield.This is minimize the eye movement of the driver, allowing him to focus on the road while driving, as well as, receive navigational assistance.

**Requirement 3: Improved Voice Commands**

**Proposed Solution: Designing a voice user experience that aligns with user's expectations**

A recent study conducted by the linguistics department at the University of Oxford revealed that humans have a typing speed of 40 words per minute but they can speak upto 150 words in the same time span.Hence, a voice assistant is definitely a better choice to optimize the driving performance, while causing minimal distractions.The success of the voice assistant also depends on others factors like the pitch of the tone, language used by the assistant, tone of the language, clarity of speech, empathy shown by the assistant etc.

The ideation phase concluded with an example ideation map, that shows a concise summary of the ideas finalised for the design prototype:

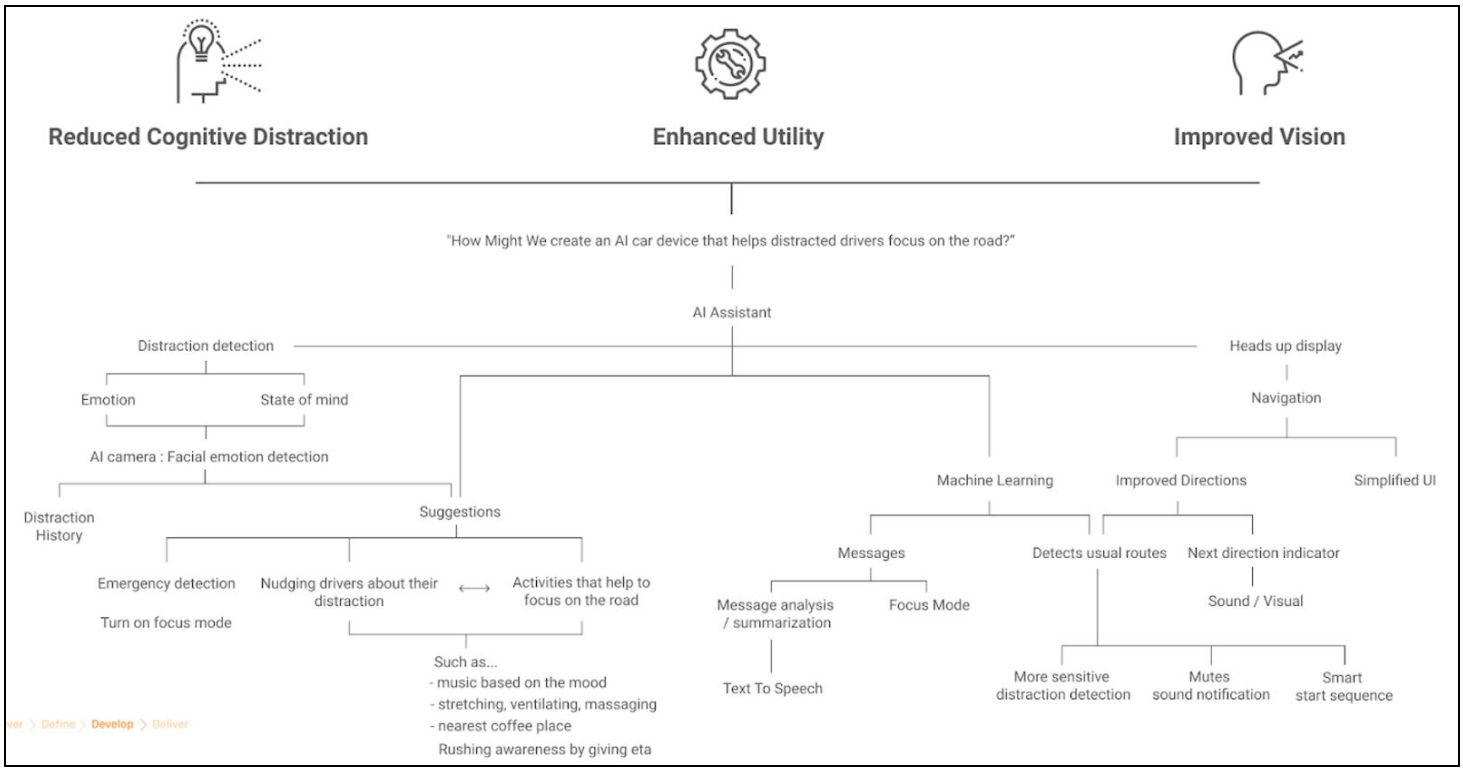


**Figure 2:** Function Map(Source: http://seungpil-lee.com, Accessed on May 16, 2020)

### 5.2 Voice Assistant Features:

The proposed voice assistant has the following three features that caters to user requirements of all the groups:

- **Personalized AI Assistant:** Proactive AI powered voice assistant that navigates the user to a better driving experience.The assistant will be wired using natural language processing techniques to read out only high priority messages, ignoring the non-important ones.It can also offer recommended replies or allow the driver to frame the reply through the voice feature, thus, saving the driver from mental load to respond to messages and preventing distraction.The assistant is also able to turn on the 'Focus mode' for the driver which silences notifications from all apps.
- **Computer Vision Enabled:**The assistant has an inbuilt camera feature that gathers visual cues about the driver's facial expressions, tone and body movements and offers solutions accordingly to minimize distraction.The assistant can nudge the driver to move to a safety location if he/she is feeling drowsy, unhealthy or emotionally disturbed. It will also keep track of the distraction history of the driver and allow him/her to view it.
- **AR powered Windshield Projector:**The in-car voice agent has a feature to project the mobile screen visuals on the windshield of the car so that the driver does not need to check

his phone repeatedly to view text messages, app notifications and Google Map navigation directions.

Given below is a sample diagram for the voice flow for emergency detection scenario.

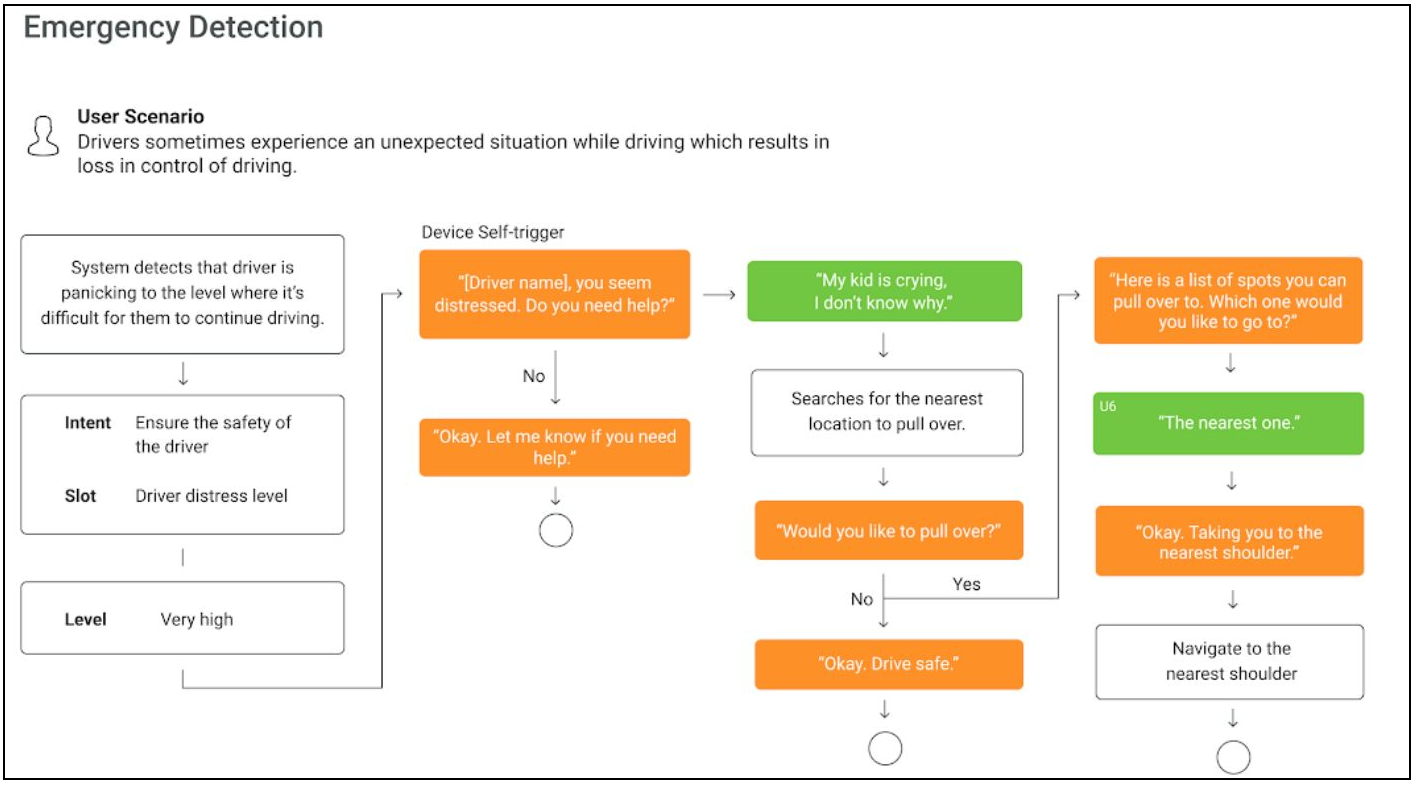


**Figure 3:** Voice Flow Chart for Emergency Detection (Source: http://seungpil-lee.com, Accessed on May 16, 2020)

## 6 Evaluation of the User Experience

The success of any technology, service or product is dependent on how well it fulfills the user's requirements and how intuitively seamless is its functionality.Being empathetic towards the user's requirements and their pain points plays a pivotal role in the user adoption of any technology. To gain user feedback about the in-car voice assistant, user testing was conducted on four participants using heuristic evaluation to figure out the usability issues in the personal assistant.The table below shows the usability test results(yellow signifies positive feedback, green signifies neutral feedback and red signifies negative feedback):

| Usability Test | Participant Feedback |
|---|---|
| Conversational Assistant's voice tone | Positive feedback in general (3 out of 4 respondents showed approval) |
| Visual Interface and Navigation | User Interface was cluttered and unclear about functionalities |
| Start Sequence | Feedback was a tie with 2 respondents having positive opinions and other 2 having negative feedback |
| Texting | All respondents said that the text writing functionality should incorporate writing lengthy text messages as well. |
| Emergency Detection | Positive feedback in general |
| Distraction Detection | Icons for different distractions are unclear in their meaning |
| Focus Mode | Notification feature is unnecessary and frustrating |
| Infotainment | Positive feedback in general |

Based on the feedback received, the following functionalities and features were modified to align them with the user's expectations:

- Increased the spacing among the UI features for clarity and inserted swipe left/right feature to accommodate functionalities on the next screen.
- Enhanced icons for intuitiveness
- Removed the notification feature from the focus mode

## 7 Limitations and Future Research

Like many other studies, this research project also has its own set of limitations. The usability tests for this study were conducted in an online mode on the participants' own devices(mobile phone, laptop, desktop). Because of this restriction, the modal quality may not have resembled the realistic experience for the respondents to respond in ways that would hold valid in real life scenarios. Thus, future usability testing should be conducted in more immersive simulators and for actual driving contexts.

The future milestones of this research include implementing an Alexa skill for the proposed prototype.The Alexa skill would actually enable the customers to use the design solution proposed in this project. Our society functions more on gender stereotypical basis, where one gender is trusted more to perform a specific task rather than their reposing equal trust on both genders.Given this biasness, it would be interesting to evaluate the user's trust based on the gender of the conversational agent.This is another future milestones that I plan to carry out for the proposed prototype.

## 8 Conclusion

With every technological advancement, it is critical to understand how broadly it is adopted.Conversational assistants have a large user base and automobiles ranks second after smartphones in terms of monthly active users and total customer reach. To maintain these figures and ensure the same rise, it is important to have human centered design approach for designing VUIs.From the above design solution, I realized that the following three principles must be essentially followed while designing unique intuitive experiences for voice:

- **Cooperative Principle:**It describes the rules and assumptions human beings abide by in our conversations and we do so entirely subconsciously.This maximizes the quality, quantity, relevance and manner, which can be used as reliable guidelines while creating conversations.
- **Turn Taking Principle:** This principle states the mechanism by which we resolve ambiguity and repair conversations in order to progress to a solution.
- **Designing for user contexts:** Talking to machines is still in its infancy and people still find it socially awkward.Therefore, the designers should look to design for simple, intuitive and high value conversations so that voice interactions have less friction than touch.

The proposed in-car voice assistant makes use of all the above mentioned three principles to offer a rewarding driving experience to the user. Since the assistant has a combination of visual and audio features, it was important to strike a balance between the touch and the audio mediums.I carefully analysed the user requirements to come up with proper choreographies for the audio and video channels.Another important factor to keep in mind is that we need to make conversations work well even when it goes off track.These can be points of great user frustrations and if handled well, they are great opportunities to rebuild user trust and confidence.Thus, all efforts have been made to minimize the points of user frustration and offer a simple yet powerful design solution for distracted driving.

## References

[1] A. B. Kocaballi, L. Laranjo and E. Coiera.(2019, Mar.). Understanding and Measuring User Experience in Conversational Interfaces. *Interacting with Computers*[Online].*31(2)*, pp. 192-207. Available: https://doi.org/10.1093/iwc/iwz015

[2] Z. Wei and J. A. Landay(2018, Jun.). Evaluating Speech-Based Smart Devices Using New Usability Heuristics. *IEEE Pervasive Computing*[Online]. *17(2)*, pp. 84-96. Available:https://hci.stanford.edu/publications/2018/speech-he/speech-heuristics-ieeepvc-20180406.pdf

[3]L. Clark, A. Ofemile and S. Adolphs(2016). A Multimodal Approach to Assessing User Experiences with Agent Helpers. *ACM Transactional Interactactive Intelligent Systems*[Online].*6(4)*, pp. 291–298. Available: https://doi.org/10.1145/2983926

[4]V. Këpuska and G. Bohouta, "Next-generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home)," in IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, 2018, pp. 99-103.

[5] J. Schmidtler, K. Bengler, F. Dimeas and A. Campeau-Lecours, "A questionnaire for the evaluation of physical assistive devices (QUEAD): Testing usability and acceptance in physical human-robot interaction," in IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, 2017, pp. 876-881.

[6]F. Bentley, C. LuVogt, M. Silverman, R. Wirasinghe, B. White and D. Lottridge, "Understanding the Long-Term Use of Smart Speaker Assistants" in Proc. ACM Interaction Mobile Wearable Ubiquitous Technology, 2018, pp. 256-264.

[7] S. Amershi, D. Weld, M.Vorvoreanu, A.Fourney, B.Nushi, P.Collisson, J.Suh, S.Iqbal, P.Bennett, K.Inkpen, and Eric Horvitz, "Guidelines for Human-AI Interaction" in Conference on Human Factors in Computing Systems Proceedings, New York, NY, 2019, pp.62-78.

[8] G. Lugano, "Virtual assistants and self-driving cars," *in 15th International Conference on ITS Telecommunications (ITST)*, Warsaw, 2017, pp. 1-5.

[9]M. F. M. Idros, A. H. A. Razak, S. Al Junid, A. K. Halim and N. Khairudin, "Capability of Voice Recognition System for Automatic Signal in Autonomous Vehicle (AV) Application," in *IEEE 5th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA)*, Songkla, Thailand, 2018, pp. 1-6.

[10] J. Park, H. Son, J. Lee and J. Choi(2019, Jan.).Driving Assistant Companion With Voice Interface Using Long Short-Term Memory Networks. *IEEE Transactions on Industrial Informatics*[Online]. *15(1)*, pp. 582-590.Available: https://ieeexplore.ieee.org/document/8423671

[11] D. Pal, C. Arpnikanondt, S. Funilkul and V. Varadarajan, "User Experience with Smart Voice Assistants: The Accent Perspective," *in 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kanpur, India, 2019, pp. 1-6.

[12] S. Han and H. Yang(2018). Understanding Adoption of Intelligent Personal Assistants: A Parasocial Relationship Perspective. *Industrial Management & Data Systems*[Online].*118(3),* pp. 618-636. Available: http://iranarze.ir/uploads/2018/02/E6014-IranArze.pdf

[13]D. L. Strayer(2017).The Smartphone and the Driver's Cognitive Workload: A Comparison of Apple Google and Microsoft's Intelligent Personal Assistants.*Canadian Journal of Experimental*

*Psychology*[Online].*71(2)*, pp. 93-110. Available: http://appliedcognition.psych.utah.edu/publications/smartphone_workload_apple_17.pdf

[14]G. McLean and K. O. Frimpong(2019). Hey Alexa ... Examine the Variables Influencing the Use of Artificial Intelligent in-Home Voice Assistants. *Computers in Human Behavior*[Online].*99*, pp. 28-37. Available: https://doi.org/10.1016/j.chb.2019.05.009

[15] http://seungpil-lee.com/project/cair-2