

THE FUTURE OF VOICE ASSISTANTS IN CARS: CURRENT GAPS AND OPPORTUNITIES TO ENHANCE USER EXPERIENCE

Ashlesha Vishnu Kadam
Seattle WA, USA

Abstract— In-car voice assistants like Google Assistant, Siri, Alexa, and others have become increasingly important modes of communication and control in the car. However, there are seldom any observable in-car user experiences with voice assistants that seem to have been built car-first, i.e. with the car use case in mind as the top one. In this article, we first examine the ways in which voice assistants are integrated into the car. We then explore the top concerns and gaps in the current voice assistant offering in the market and a framework for designing thoughtful user experiences in the car. Finally, this article provides a brief outlook about the future of voice assistant adoption and usage in the car.

Keywords— ASR, NLU, NLP, voice assistant

I. INTRODUCTION

There is no doubting that Voice Assistants (VAs) are ubiquitous. While Voice Assistants lend a natural, intuitive and simple way to accomplish tasks using just your voice, the comfort, convenience and safety of being able to perform these tasks on the road without having to take your hands off the wheel is unparalleled. An estimated 125MM customers in the US alone use voice control technology in the car during their average daily 27.6-minute commute. This adoption, coupled with the increase in consumption of content due to “return to commute” post pandemic, indicates that voice technology is a growing opportunity.

II. CURRENT VOICE TECHNOLOGY IN THE CAR

Before diving into the voice assistant experience in the car, it is important to understand how a voice assistant works end-to-end. See Fig. 1 as a reference. Assume a hypothetical voice assistant Nova. When the user mentions the invocation word “Nova”, the voice assistant Nova starts listening to the voice signal that follows from the user. This voice signal is then translated into a string of tokens in the ASR (Automatic Speech Recognition stage). The string is then processed to make semantic sense out of it. For example, here the request is understood to be a request for

music that has released recently from the hip-hop genre. This interpretation is then passed on to the back-end for performing execution to fulfil the request. The query is then fulfilled, appropriately decorated and returned to the user in the form of a TTS and/or additional content (e.g. here, playback of tracks).

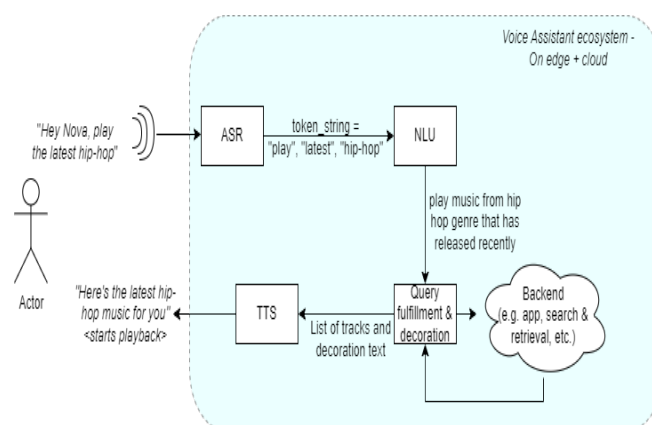


Fig. 1. Overview of end-to-end working of hypothetical voice assistant Nova (Kadam, 2023)

While customers have discovered and adopted the convenient experience that Voice Assistants bring to them in the car, VAs are still used in the car for arguably mundane tasks like navigation instructions to the nearest gas station, calling hands-free, playing music or audiobooks or podcasts, getting the weather forecast, and so on. On one hand there are smartphones offering familiar voice technology (try “Hey Google, where’s the nearest gas station?”), and on the other, there are in-built voice assistants like Ford offering Amazon’s built-in Alexa, Chevrolet, Renault providing hands-free tech via Android Automotive OS and Tesla offering its own voice technology. See Fig. 2. for an overview of the two options in which automobile companies might choose to integrate with voice assistants.

As seen in Fig. 2., the path from A1 through A5 demonstrates the end-to-end flow of a 3P voice assistant implementation (e.g. using a car device that Google specifically creates for the car, or Amazon creates for Alexa), while the path B1 through B5 demonstrates the integration with a voice assistant that is either native (1P – green arrows) or simply a native skin over an existing 3P voice assistant (3P – blue arrows). An example

of the former is Tesla’s implementation of its voice assistant that is purely native and does not leverage any existing voice assistants like Alexa, Google Assistant or Siri. An example of the latter could be a hypothetical automobile company called Space whose cars have a voice assistant called “Hey Rocket” but is really invoking Alexa in the backend.

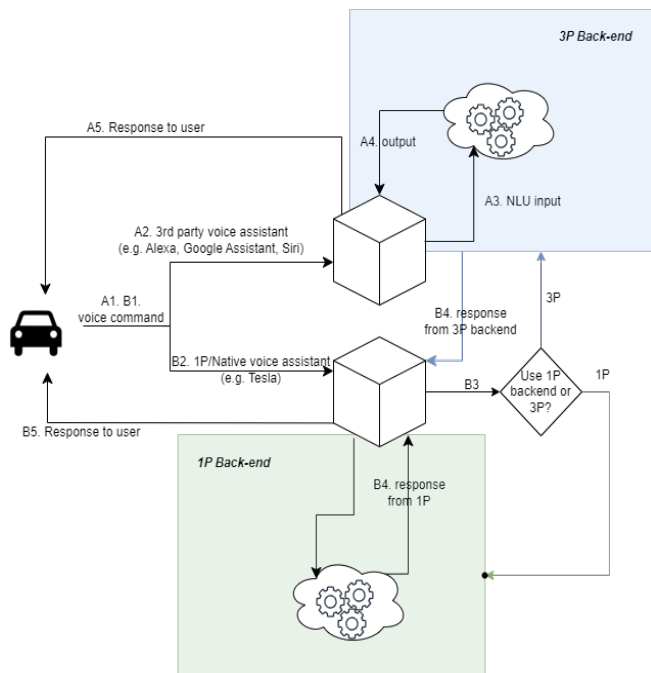


Fig. 2. 1P and 3P voice assistant integration options for automobiles

III. CONCERNS AND GAPS IN THE CURRENT VOICE TECHNOLOGY SOLUTIONS IN THE CAR

Despite voice technology being available in the car, there are shortcomings in the voice experience inside the car. A few include:

3.1 Safety Concerns

While developers of voice technology claim that using voice assistants in the car is safer than having the driver fiddle with their phone while driving, there are research reports that suggest that use of voice assistants in the car can be a powerful and lingering distraction. In fact, the use of voice in the car has been likened to “balancing a checkbook while driving”.

3.2 Better voice understanding

Voice assistants continue to fall short of customers’ expectations when it comes to how they interpret what a user is asking for. For example, voice assistants expect questions to be framed in a certain way that is understandable to the voice technology, don’t do as well when it comes to understanding users with speech imperfections like lisps, different accents, multilingual style of interaction, lack of awareness of context,

and more. These situations are further exacerbated in the car where the surrounding is noisier, increasing the chances of failure in speech recognition.

3.3 Speed, Accuracy and Responsiveness

Low latency is a key factor to making a good voice assistant user experience. Particularly in the car, poor latency can make it hard for voice assistants to provide a delightful experience. If a customer wants to quickly know whether or not they need to make the upcoming right turn on the signal, if they ask their voice assistant this question, they expect a quick, to-the-point and accurate response. However, based on anecdotal observations, customers often end up having to wait for their question to go through for a few precious seconds, only to have the voice assistant fail at understanding the command or worse still, providing an answer to an incorrectly heard and understood question. These poor experiences deter users from re-using these voice assistants.

IV. DESIGNING THOUGHTFUL VOICE EXPERIENCES IN THE CAR

A fundamental reason for voice-forward experiences in the car not being top notch could be that these voice experiences have often not been crafted in a bespoke fashion for the in-car environment. To craft the ideal experience, one needs to step back, step into the car, and think about how one would build a voice assistant specifically for the in-car experience that would not just address the current gaps and concerns, but also meet, or even exceed the expectations of the user in the car. In this article, we propose a modification of the Kano framework and apply it to the use case of voice assistants in the car. The requirements have been gathered via literature review done across the articles mentioned in the References section. See Fig. 3 for reference of this framework. We characterize the (1) Must-be requirements as “Nailing the Basics / Must-haves”. These are implied and obvious requirements that have to be met for a voice assistant to even be considered for usage in the car, (2) One-dimensional requirements as “Enhancing the Everyday Experience / Performance”. These are requirements that have been specified and articulated explicitly. The more these requirements are met, the more the perceived performance of the offering, the more customers will enjoy the experience, and (3) Attractive requirements as “Going Above and Beyond / Delighters”. These are requirements that are not expressed by the customer but if encountered would pleasantly surprise and delight customers.

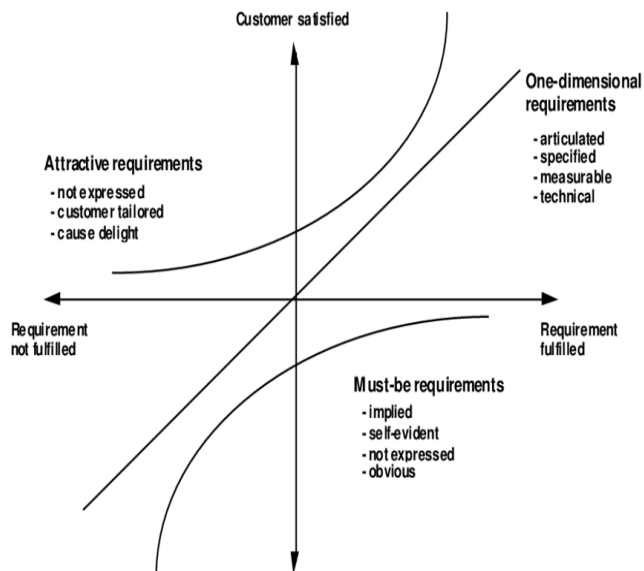


Fig. 3. Kano's model of customer satisfaction Error! Bookmark not defined. (Berger et al., 1993)

We look at the application of this framework in 3 steps:

4.1 Nailing the basics / Must-haves

These are the must-haves that, if incorporated, would address the key concerns that keep more users from adopting voice technology while driving

4.1.1 Serve basic command - control interactions

Drivers think of using voice assistants in their car for activities that are important while driving like knowing what the weather forecast is, being able to call someone, knowing how the traffic's like, how much fuel's in the tank, how far the nearest gas station or bathroom is. At a minimum, for users to even think of using voice assistants for these tasks, it is critical that these basic needs are well served. If a user has to repeat their request 4-5 times for the voice assistant to understand, or if the voice assistant answers something irrelevant because it did not hear or understand the command, there's little chance that the user would give the voice assistant another chance. Speech recognition also needs to be high quality in the car in order to be able to suppress ambient noise and understand the driver's voice clearly.

4.1.2 Ensure Safety First

Arguably, the most important aspect to keep in mind while designing voice experiences in the car is to make sure they help drivers accomplish tasks hands-free Error! Bookmark not defined. instead of distracting them. Thoughtful designing of experiences, like a shorter verbal response, or acknowledgment of a command, can go a long way in minimizing distraction and reducing the driver's temptation to look at the voice assistant to see if they heard them or not. For example, when a user asks the voice assistant what the weather is like, instead of a long-winded response like "Today

in San Diego the high is 79-degree F and the low is 58-degree F. You can expect a sunny weather with a 50% chance of mild showers. Don't forget your umbrella!" the response should simply be "It's warm right now at 72 degrees F". It is also important to understand the context of a request because a driving customer might not always articulate comprehensively. For example, if the driver just says "How far is it?", the voice assistant shouldn't ask "What destination are you referring to?". Instead, the voice assistant should remember that just a minute ago, the driver asked about the nearest coffee shop, so they must be referring to that coffee shop. The response could simply be "The nearest Starbucks is 2 miles out north".

4.1.3 Embody ideal Form factor

It is also important to be deliberate about the form factor of the voice assistant. To avoid having a visually distracting object, it is better to have a minimal, preferably screen-less, design of the voice assistant. If a multi-modal voice assistant is desired, say to allow the ability to display navigation, infotainment pages, etc. then it is important to think about both visual and audio modalities to craft the perfect experience. For example, there shouldn't be any change to the screen when not needed (e.g. for when you just ask for the time), but if you ask for navigation instructions to the nearest burger place and there are 2 possible routes, it might be easier to show those two options on the screen than for the voice assistant to verbally explain the two route options.

4.2. Enhancing the everyday experience / Performance

These aren't exactly must-haves, but incorporating these into the experience that you craft for your users in the car will go a long way in helping with the adoption of voice assistants as a daily habit.

4.2.1 Keep Driver Entertained

Listening to the news, music, podcasts, audiobooks, and more is one of the top things that drivers resort to keep boredom at bay during long drives. Instead of having the driver fiddle with their phones, voice assistants are perfectly poised to do these tasks at command. Users should be able to summon the voice assistant to jump into the music they were listening to, switch it up to something upbeat or relaxed, resume their audiobook, and more using just their voice.

4.2.2 Understand Natural Language

High quality speech recognition is a must-have in the car where a user can't afford to repeat the same thing multiple times while focusing on driving. At the same time, speed, accuracy and responsiveness of the voice assistant becomes important for a performant experience. With advancements in compute technology / IoT, processing user requests on the edge can make a big difference in perceived performance of the voice assistant.

4.2.3 Keep Driver awake

We have different technologies that can detect sleepiness in a driver from their speech rate, driving pattern, and other signals. Voice assistants could play the role of a ‘companion’ who is checking in periodically to make sure the driver isn’t dozing off. Voice assistants could also simply keep polling the driver every now and then, especially through late night drives, to ward off their drowsiness.

4.2.4 Adapt TTS

The Text-to-Speech (TTS) response of a voice assistant to a command should be modulated in pitch, tone and brevity based on what is known about the driving circumstances. If there is high traffic density, it might be preferable to have quick short acknowledgments by the voice assistants whereas under regular circumstances, TTS can be the usual one.

4.3 Going above and beyond / Delighters

These are most certainly not essential to have, but when available, can delight the user and create positive associations with the voice assistant so that they keep coming back for more.

4.3.1 Provide Multi-person Experiences

Knowing if there is just the driver or their family or friends in the car can help the voice assistant tune the experience for the right audience. What music recommendations are chosen for playback, what news is relayed and even what recommendations for the nearest lunch place are served can change based on who exactly is in the car to enhance personalization and the degree of user satisfaction. For example, if a driver is driving alone to office on a Monday morning, they might like some upbeat hip hop. But when the same person is driving his young family to the nearest park, they might like the latest hits by Kidz Bop Kids.

4.3.2 Adapt Language

If it is detected that the driver is struggling to respond in the default language of the voice assistant, the voice assistant should smartly switch to the language the driver is most comfortable in so that it isn’t a cognitive load on the driver to use the voice assistant and they feel more comfortable and understood. Or if there’s a multi-lingual group traveling in the car, then the voice assistant should switch its language as well as recommendations for news, music and more to a language that works the best for the group instead of just the driver.

4.3.3 Keep Interactions Kid friendly

Knowing and tuning experiences to make them kid-friendly could be a huge delighter for families commuting in the car. For example, automatically playing only the clean version of songs, not playing inappropriate ads, highlighting PG content before showing it, and more can make voice assistants a compelling companion for families in the car.

V. LOOKING FORWARD

The auto industry is at an inflection point where in a few years, the mix of cars that are getting bought might skew more and more towards autonomous vehicles and electric vehicles instead of the incumbent cars. This could have implications on the use of voice assistants in the car.

A. Influence of Autonomous Vehicles

The shift in the auto industry towards EVs and autonomous vehicles could imply even more time for drivers to engage with voice assistants. As EVs and autonomous vehicles become more commonplace, the appeal of voice as an interface to interact with the world from the car will increase even more.

B. Law, Ethics and Voice Assistants

There aren’t specific laws or corporate governance when it comes to usage of voice assistants in the car. There are also some meaty unresolved questions in the ethics space, like who is to blame if a driver gets distracted while talking to a voice assistant, causing an accident. As voice assistants become ubiquitous in the car, some of these regulations coming in might change the way voice assistants are, or can be, used in the car.

C. Augmented Reality

With augmented reality-based experiences entering the car to make it a safer, smoother and delightful experience for users in the car, voice assistants might be a complimentary medium to provide a holistic improved experience. We are barely scratching the surface in the space of AR and voice technology together in the car.

VI. CONCLUSION

The outlook is bright. Voice assistants continue to remain a highly desirable medium of interaction in the car, and the ease, convenience and intuitiveness that voice assistants offer fit perfectly with the needs of the user in the car. Designing thoughtful car-first experiences for voice assistants can go a long way in increasing the adoption of voice assistants and making them a daily habit for users in the car.

VII. REFERENCES

- [1]. Singh, Deepika & Kaur, Gurwinder & Bansal, Ankit. (2019). How Voice Assistants are Taking Over Our Lives – A Review, JETIR May 2019, Volume 6, Issue 5. ISSN-2349-5162
- [2]. <https://voicebot.ai/2022/04/07/in-car-voice-assistants-reclaim-some-active-users-as-pandemic-subsides-new-report/>
- [3]. <https://www.census.gov/newsroom/press-releases/2021/one-way-travel-time-to-work-rises.html>



- [4]. https://www.insideradio.com/free/third-of-americans-spending-more-time-with-digital-audio-post-pandemic-survey-finds/article_989334b4-9af7-11ed-a0fa-5b18f9aa884c.html
- [5]. <https://www.soundhound.com/blog/how-drivers-are-actually-using-voice-assistants-in-cars-bb1e3972301b/>
- [6]. <https://www.ford.com/alexa/>
- [7]. <https://blog.google/products/android/android-auto-new-features-google-io-2023/>
- [8]. https://www.tesla.com/ownersmanual/model3/en_us/GUID-EA1715B0-A3A6-454E-995A-8AA2C3A32D44.html
- [9]. <https://voicebot.ai/2019/06/17/study-using-voice-assistants-in-cars-as-distracting-as-drunk-driving/>
- [10]. Strayer, D.L., Cooper, J.M., Turrill, J., Coleman, J.R. & Hopman, R.J. (2015). The Smartphone and the Driver's Cognitive Workload: A Comparison of Apple, Google, and Microsoft's Intelligent Personal Assistants (Technical Report). Washington, D.C.: AAA Foundation for Traffic Safety.
- [11]. <https://www.nytimes.com/2015/10/22/science/cars-voice-activated-systems-distract-drivers-study-finds.html>
- [12]. <https://blog.webex.com/engineering/robust-nlp-for-voice-assistants/>
- [13]. <https://levelup.gitconnected.com/how-nlp-improves-multilingual-text-to-speech-voice-assistants-practical-case-inside-91893f0d04d>
- [14]. Bogdan, Răzvan, Alin Tatu, Mihaela Marcella Crisan-Vida, Mircea Popa, and Lăcrămioara Stoicu-Tivadar. (2021). A Practical Experience on the Amazon Alexa Integration in Smart Offices, *Sensors* 21, no. 3: 734. <https://doi.org/10.3390/s21030734>
- [15]. Sauerwein, Elmar & Bailom, Franz & Matzler, Kurt & Hinterhuber, Hans. (1996). The Kano Model: How to Delight Your Customers. *International Working Seminar on Production Economics*.
- [16]. <https://www.soundhound.com/blog/how-drivers-are-actually-using-voice-assistants-in-cars-bb1e3972301b/>
- [17]. Emily Buteau & Joonghwa Lee (2021) Hey Alexa, why do we use voice assistants? The driving factors of voice assistant technology use, *Communication Research Reports*, 38:5, 336-345, DOI: 10.1080/08824096.2021.1980380
- [18]. Hu Q, Mohamed T, Xiao W, Ma X, Gao X, Gao Z, Arava R, Abdel Hady M. Frost. (2022): Fallback Voice Apps Recommendation for Unhandled Voice Commands in Intelligent Personal Assistants. *Front Big Data*. 2022 Apr 25;5:867251. doi: 10.3389/fdata.2022.867251. PMID: 35547192; PMCID: PMC9084426.
- [19]. Md. Ebrahim Shaik, (2023) A systematic review on detection and prediction of driver drowsiness, *Transportation Research Interdisciplinary Perspectives*, Volume 21, 2023, 100864, ISSN 2590-1982, <https://doi.org/10.1016/j.trip.2023.100864>
- [20]. <https://uxdesign.cc/voice-user-experience-design-and-prototyping-for-mere-mortals-ef080c843640>
- [21]. <https://venturebeat.com/datadecisionmakers/why-voice-assistants-could-be-the-future-of-user-experience/>
- [22]. <https://summalinguae.com/language-technology/language-support-voice-assistants-compared/>
- [23]. Pal, Debajyoti & Arpnikanondt, Chonlameth & Funilkul, Suree & Varadarajan, Vijayakumar. (2019). User Experience with Smart Voice Assistants: The Accent Perspective. 1-6. 10.1109/ICCCNT45670.2019.8944754.
- [24]. Ashlesha Vishnu Kadam, Designing Thoughtful Experiences for Kids on Voice Assistants, *International Journal of Artificial Intelligence & Machine Learning (IJAIML)*, 2(1), 2023, pp. 75-81 doi: <https://doi.org/10.17605/OSF.IO/HKTS8>
- [25]. <https://www.caranddriver.com/news/a39998609/percentage-of-electric-cars-usa/>