A New Race for Luxury Cars: How to Implement Vocal In-car?

***

By Isabelle ROUSSEAU

Under the direction of Anne-Flore MAMAN LARRAUFIE (Ph.D.) and formatted by Cécilia GOUTRAN

info@semioconsult.com

Even if vocal recognition is not (yet?) the dominant design, vocal assistants are part of our daily lives. Could vocal be a good answer to the increasing number of features embedded in luxury cars? Will vocal features be a criterion for luxury vehicle purchase? What technical and security aspects should be considered? Following a previous article[1] on *Driving Innovation in the Car Industry: The Case of Luxury Cars*, we will unveil the lessons of the Delphi study conducted on the use of voice in luxury cars and the way to implement such strategy by analysing and understanding better their clients and eventually improving the entire driving experience. The study was supported by a panel of experts located on four continents, of different age and occupation.

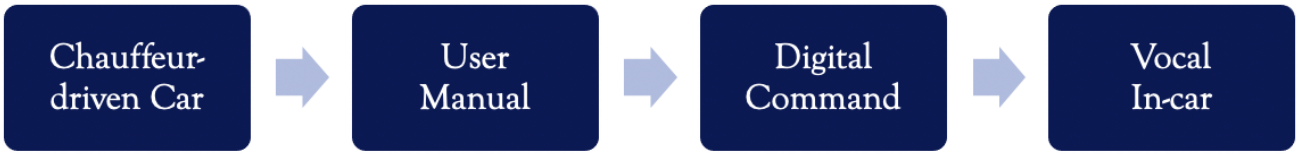
As part of her thesis at ESSEC, Isabelle Rousseau conducts a study **upon vocal in-car technology applied to luxury car industry based on what deeply define luxury and on the way luxury companies may innovate.**

Acknowledging the mutation of the whole automotive industry, SémioConsult® supports Isabelle Rousseau's proposal to try to understand how to drive innovation through luxury cars. We supervised her thesis and guided her in the Delphi method, one of our expertise.

Do not hesitate to get back to us, should you need a deeper analysis of the market or should you have any queries regarding this project or Delphi studies: info@semioconsult.com.

[1] https://www.semioconsult.com/wp-content/uploads/2021/02/2021-01-Article-Driving-Innovation-through-Luxury-Cars.pdf

EVOLUTION

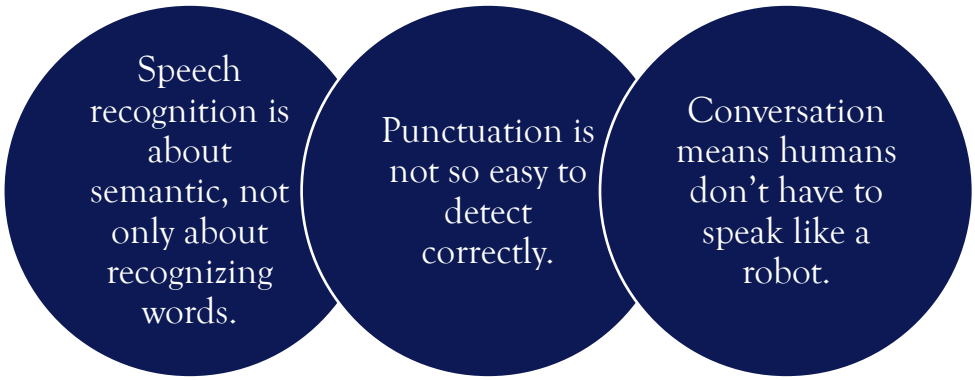


Vocal assistants are part of our daily life, via your smartphone (Hey SIRI, or Hey Cortana) or via a speaker "Ok Google". Voice User Interfaces (VUI) enable communication between humans and machines through voice-based interactions.

How does it work?

Speech recognition is a complex subject. Converting natural language into text mobilizes knowledge in linguistics, acoustics, semantics, mathematics. The system catches the sound of your voice, predicts the words that could be associated and defines the meaning of your sentence in a quick way in order to leave space for a potential dialogue. Difficulties may come from the way humans speak with regional dialects, different speeds, the existence of surrounding noises, but also the number of people speaking. Furthermore, expressions like "ahhs" or "hmmm" may be misunderstood and have to be isolated from the words recognition since they are more likely to express hesitation or the activity of thinking. Finally, each language has its own particularities, thus translation of the first language model is often not enough. As an example, Google has announced that its machine learning algorithms have now achieved a 95%-word accuracy rate for the English language but has to improve on other languages.

Three major points have to be taken in consideration:



1. **Speech recognition is about semantic, not only about recognizing words.** The difference between the Natural Language Processing (NLP) and the Natural Language Understanding (NLU) is that NLP

interprets word to word what humans say, whereas NLU by adding an Artificial Intelligence's feature identifies the profound meaning of what is said to determine intentions.

- ⇨ For example, if you say *"I am looking for an Italian restaurant except a pizzeria"*, word recognition will understand "Italian", "restaurant" and "pizzeria" and not "except pizzeria".

2. **Punctuation is not so easy to detect correctly**. Punctuation gives meaning to the sentence and if not taken into account it may be a source of misinterpretation. Think about:
   - ⇨ *"I love baking my family and my friend."* vs *"I love baking, my family, and my friends."*
3. **Conversation means humans don't have to speak like a robot.** Conversational quality will come from functional and technical abilities. Humans want to talk to machines just the way they talk to each other, that is to say in "natural" language, not in a machine language.

What's new?

Big Data and growing computer power have significantly improved the accuracy of voice technology. Vast amounts of unstructured records in various languages and accents, noisy or calm environment are now massively collected and used to train the different speech recognition solutions. The more advanced solutions integrate grammar, syntax, structure but also use machine learning and Artificial Neural Networks (ANNs) to understand human speech in long and complex sentences - after extensive training.

Almost as efficient as humans

The goal in speech recognition is to reach and exceed human capacity. Recently, IBM has claimed to reach a word error rate (WER) of 5.5% in English[2] with its Watson Assitant solution and the French company Allomedia has claimed[3] an average WER of 4.1% on French conversations, a result better than humans stuck at 5% to 7.5%, and even better at specific expressions.

Sentiment and emotion

A next step will be sentiment analysis to interact better with the vocal machine; not only to adapt the conversation but also to personalize on a daily basis the available services. Mood and emotion influence the way

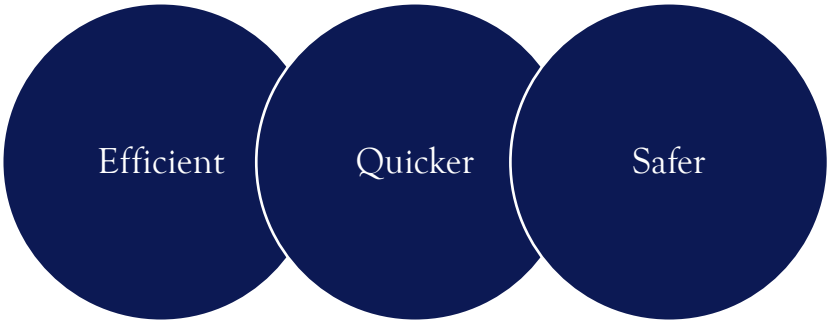
[2] https://www.ibm.com/topics/speech-recognition
[3] https://www.allo-media.net/en/expertise/

we speak.

A Toyota's prototype named "Mobility Teammate Concept"[4] was for instance studied, after facial and voice analysis, to activate autonomous driving in case of some stressful mood being detected.

Some professionals think about not only integrating emotion's recognition but also giving voice assistant a personality and emotions in order to be more natural in our daily life, and relying on anthropomorphism[5] (Borelle, Velkovska, & Zouinar, 2018). This phenomenon may impact the interaction between machine and human.

Use cases and benefits



Efficient and quicker

In 2019, 70% of people were satisfied by using their voice-based speaker device like Google Home or Amazon Echo[6]. What about drivers? Due to the increase of services on board, in-car voice assistant has become part of the driving experience. Total users and monthly active users of voice assistants are now more numerous in cars than on smart speakers[7], and the increase of services on board; conversational assistants tend to go mainstream.

Efficiency is real if the used interface is fully vocal and very well designed. As an example, a single query may avoid step-by-step protocols. Let's consider this sentence: "I want to modify my ticket for tomorrow at 10pm" is quicker than answering to questions like: which day ? at what time? …

Safer

In 2017, only-voice command systems like MyLink were tested fluider and less demanding to the driver than the Connect system. In 2018, Apple CarPlay and Android Auto from Google were both launched with the objective of minimizing the driver's distraction. A lot of activities may distract the driver from concentrating on his or her

[4] 2017 Article Biometrics and AI Help New Toyota Concept-i Vehicles Understand Your Emotions
[5] https://digital-society-forum.orange.com/fr/les-actus/1173-machines-parlantes-34intelligentes34-la-realite-est-bien-en-deca-des-discours-marketing
[6] https://www.capgemini.com/us-en/news/smart-talk-how-organizations-and-consumers-are-embracing-voice-and-chat-assistants/
[7] *Report in_car_voice_assistant_consumer_adoption_report_2020_voicebot*

driving. It is a well-known fact that using a cellphone while driving increases crash risk[8].

According to the same survey led by Insurance Institute for Highway Safety, manipulating in-car systems like screen and buttons is the 6th cause of distraction responsible for accidents. Voice recognition systems would allow the driver to access commands precisely without removing his or her eyes from the road.

In a study conducted by Ian Reagan, an IIHS[9] senior research scientist, analyzing glances to the road ahead (through video footage and vehicle performance data), reveals that voice-command systems do not demand as much attention as visual-manual interfaces[10], moreover this result is verified for both young and older drivers.

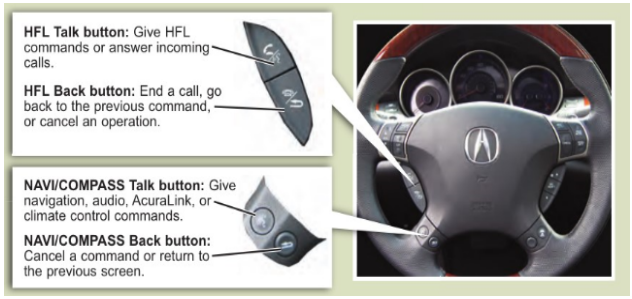
Cars.com illustration by Paul Dolan

Customers' expectations

A need for real conversation

At first vocal command experiences were deceptive. In 2009, Acura released a car in which you had to press the talk button before talking and where the exchange was based on words and not on real sentences.

[8] https://www.iihs.org/news/detail/driver-cellphone-interactions-increase-57-percent
[9] The Insurance Institute for Highway Safety (IIHS) is an independent, nonprofit scientific and educational organization dedicated to reducing the losses - deaths, injuries and property damage - from motor vehicle crashes.
[10] https://www.iihs.org/news/detail/voice-command-systems-help-older-drivers-focus-on-the-road



Source: Acura driver's guide

Despite the huge increase of devices in cars, vocal recognition is still not the dominant design. Of course current voice assistants can execute a command, respond to a user, interact with other machines but generally, these solutions force humans to adopt the phrasing of robots. The command-based dialogue should be replaced by natural conversation between the driver and his or her vehicle. The latest advances of speech recognition on long sentences may offer new possibilities and renew conversation flows.

Seamless & personalized experience

Some say that cars evolve into smartphones on wheels. Voice-based assistants are extremely popular in current digital lifestyles and consumers expect them in their interaction with their cars as well[11] (Abuelsamid, 2019).

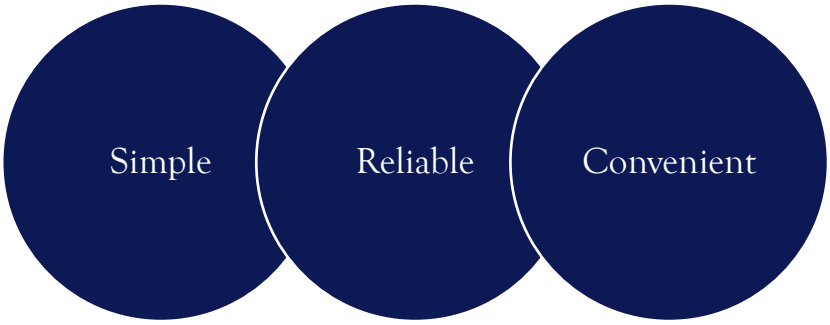
According to the 2018 Frost & Sullivan report, more than 40% drivers currently use digital assistants. Their main tasks in a vehicle are telephony, navigation and online music streaming. Instead of scrolling to choose a destination or a radio station, the driver could speak.

Integrating voice technology in a driving style does not only include driving controls but also the full digital universe or lifestyle. In other words, drivers want to keep using their digital devices and applications while driving. Separating systems will not be accepted; not only for comfort, but also to enjoy the best of both worlds. As an example, the system should interpret the integrated calendar and the navigation to choose when to leave. Another illustration would be to associate the vehicle with the home devices to offer if needed a complete and personalized experience.

[11] https://www.automotiveworld.com/articles/digital-voice-assistants-are-the-future-of-in-vehicle-control/

Background conditions of success

In the Innovator's Dilemma[12], Clayton Christensen writes that "The vehicle must be simple, reliable and convenient". Let's follow the same advice for the vocal in-car solutions.



Simple:

- Pay attention to the fluidity of the conversation. No user guides and banish the irritants. Do prefer more intuitive use cases without a long learning phase. Take an example : automatic light in tunnels is a simple application[13].

Reliable:

- Privacy by design: since voice assistants may listen to private conversations, privacy and security are paramount.
- Security: since vocal in-car assistants may influence the driving experience, security is key. Decoupling driving features from infotainment features to eliminate the risk of remote vehicle piracy.
- Accents: there are thousands of ways to say the same thing, even great providers struggled with southern and mid-western in the USA, Scottish, Indians accents, …
- Languages: it could be useful to develop multi-lingual inputs for example to improve the recognition on pronunciation of foreign expressions such as names and addresses, cities.

Convenient:

- Don't build a gadget: vocal in-car system has to answer to real use cases compatible in a luxury car environment. Think globally while allowing for regional customization.
- Agnostic: technologically speaking, the team has to be agnostic in order to be able to move across providers. For example, it is highly recommended to use neutral APIs[14] to integrate with the provider's solutions. This will be especially important to follow the improvement in semantic in each language and

[12] CHRISTENSEN Clayton (1997) The Innovator's dilemma - Harvard Business School Press.
[13] MIDLER Christophe, MANIAK Rémi, BEAUME Romain (2012) Réenchanter l'industrie par l'innovation [Re-engaging industry through innovation] – Dunod.
[14] Application Programming Interface

be useful to deploy new services in various geographical markets in order to take into account for this variability between people, geographies, and devices.

Finally, find the right balance between technology and the values of the luxury brand. The brand comes first. Be unique: in-car voice will be part of your brand. Be sure to lock careful by contract the use of voice with the supplier, so that your voice will no be use for others brands.

Insights from the Delphi Study

The purpose of this study is to obtain a reliable response to our question from a group of experts on the concept of vocal in-car in luxury cars. The Delphi study method has been designed in 1950 by Olaf Helmer at the Rand Corporation. It aims at obtaining a consensus from a group of experts through an iterative process on a topic "of the future" or "not operational".

Study protocol

To avoid influence between panelists, the principle of the study is that experts do not know each other and cannot communicate with each other. The study takes place over several rounds via electronic questionnaires.

On the first round, a questionnaire is sent to the panel of experts with precise questions. At the end of each round, answers are analyzed and returned in a synthetic form indicating the average of the responses. In the next round, the averages are shared with each respondent. They are invited if they wish, to revise their previous judgments.

After each round and some mathematical analysis of the results, some proposals are kept, and others deleted.

Sample of experts

The main criteria for selecting experts for this study is that they should all have knowledge of luxury cars professionally or personally or have carried out studies on innovation and luxury.

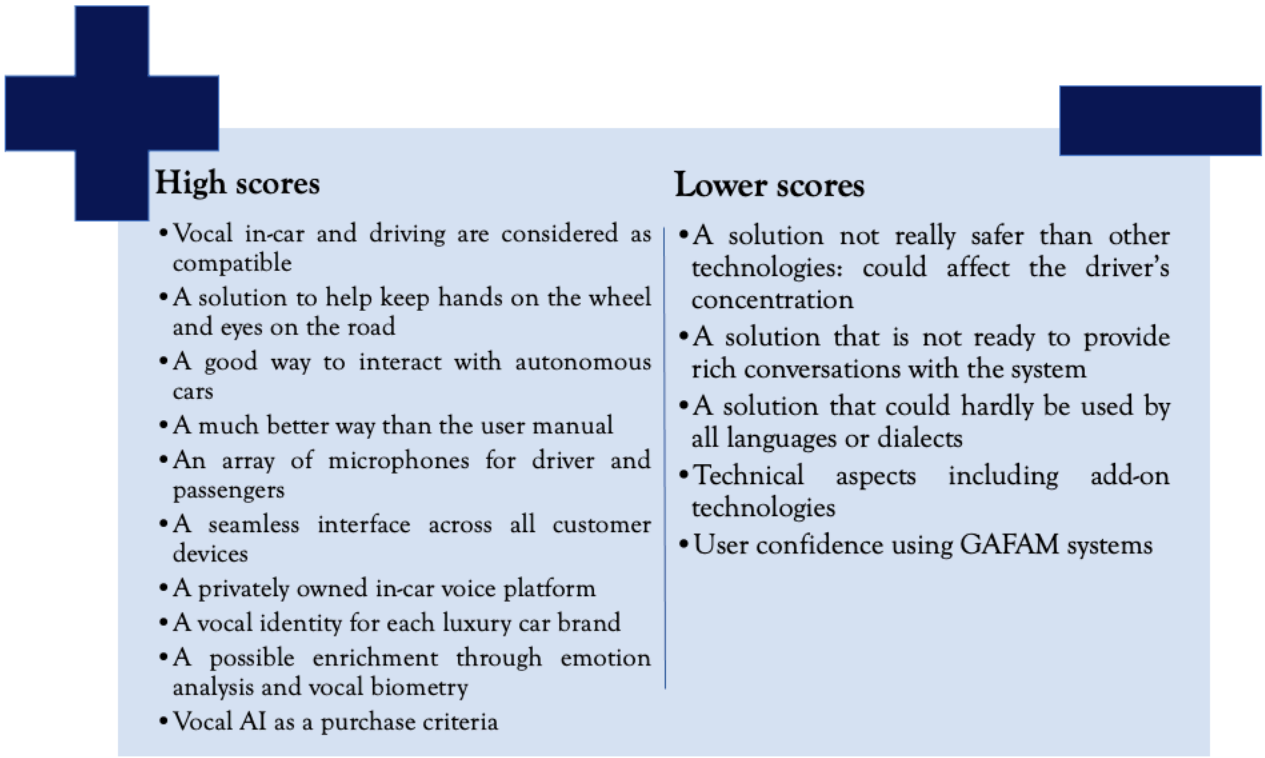
25 experts agreed to participate in this study and were then invited to complete 26 questions on issues related to the chosen topic.

Located on four continents (America, Africa, Asia and Europe), the panel was composed of different genders and generation persons like:

- luxury car collectors, amateur drivers,
- luxury carmakers professionals, in all positions: CEO, Head of Design, Human Resources,
- research professors,
- students in automotive engineering school,
- journalists in luxury car magazine.

Results and teaching

Consensus was reached on many subjects and each topic was then assigned a specific score aiming at representing the global stake behind it when talking about luxury cars development.



1. First lesson of this study: **vocal in-car and driving are considered as compatible**. Not only using voice-command will help keeping hands on the wheel and eyes on the road but it will be also a good way to interact with autonomous cars when needed. If driving with vocal command is [not really] safer speaking will [not] affect badly the driver concentration. The only point on which experts are not (yet?) convinced is on having a rich conversation with the in-car voice system to prevent from falling asleep.

   Interacting with the car in natural language is easier than consulting a growing user guide on screen. The third highest consensus is on the user manual. No need to stop driving to manipulate a screen. Based on the FAQs and chapters of the user manual, it is possible to set up a real conversation in order to transpose the intentions and wishes of the driver into voice mode. However, one does not express

oneself in the same way in writing as in speaking, as a consequence, word recognition will not be a satisfying experience, only a real conversation will.

2. Secondly the success of vocal interactions with the car depends on the quality of the vocal in-car solution starting with the quality of the microphones in such a particular environment. Using vocal in-car is so obvious that the group answered massively positive to have an array of microphones so that all – **drivers and passengers** – may benefit from vocal command, on condition that the voice of the driver takes precedence over the passengers.

   The vocal system shall be configurated to postpone the release of certain non-emergency information when the driver's concentration has to be maximum. For example: it may not be useful to push an appointment reminder at 130 km/h.

3. **Functionalities related to the time spent by the owner of a luxury car seem to be attractive for vocal in-car commands.** For instance, booking a parking place is acclaimed just like adjusting mirror, air conditioning or windows, choosing a song or a podcast.

   In the same way, sending or reading text messages rather than viewing it on a screen is approved. That means that luxury carmakers have to **offer a seamless interface** across all their customer devices, to take into account their customer's ecosystem who will enjoy to use their smartphone to send text or read a message rather than viewing it on a screen or to control home devices, as an example, to set home temperature before arriving.

4. **Functionalities directly related to driving** a luxury car are privileged over the functionalities considered as outside the vehicle. As an example, booking a ticket, or a restaurant, scheduling an appointment or searching for a product seem less attractive.

5. Moderate consensus is essentially reached on **technical aspects** and add-on technologies based on voice. The use of data by GAFAM constitutes a first obstacle to user confidence, and therefore to the acceptance of these systems. Luxury carmakers may develop their own in-car voice platform is not acclaimed because the path to develop natural language understanding (NLU) is hard, especially in more than a language. This low score is also due to customers who are now used to interact with their

own devices in natural language and want a seamless interface across all of their devices (smartphones, speakers) including the car and will expect the same experience inside their car. Luxury carmakers may use existing voice platforms as a component of their own vocal solution, and still paying a particular attention on security and privacy and preserving the brand universe.

6. It is interesting to notice that **it is more important for a luxury car brand to define the voice's identity** (like voice, tone, vocabulary) of each car than to customize the catchphrase. We can imagine two reasons: customizing the wake-up word is seen as a gadget or a way to preserve the brand identity. Another explanation could be to let the possibility to book this option to the bespoke department or limited edition that will one day propose to their customers to customize the voice by recording or buying special voices.

7. To a lesser degree, experts are convinced on the fact that vocal in-car may be enriched with other technologies such as emotion analysis, vocal biometry to open the door, to start the car or to immobilize the car if the driver is not recognized. The high level of technical requirements and privacy reasons could be a brake on immediate implementation.

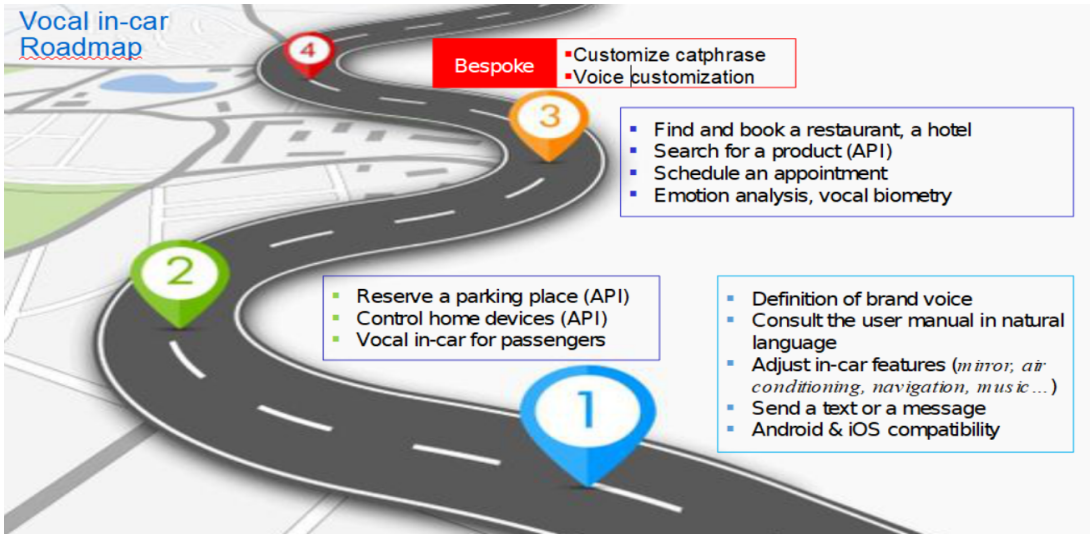
8. Finally on marketing aspects, Vocal AI is seen as a criteria for luxury vehicle purchase starting with businesswomen or businessmen who will be interested in vocal in-car, a bit higher than young people.

Conclusion of the Study

Obtained from a diverse panel in terms of occupations, ages, functions and continents, this study makes it possible to **draw a first trend on voice and automobile**. In order to increase the pleasure of traveling, luxury cars have acquired more and more new features. Vocal in-car may allow **a higher level of customization**, because it will no longer be necessary to add buttons or expand screens but only to add a catalog of intentions responding to new use cases. Keeping eyes on the road and hands on the wheel is the main advantage of the vocal in-car technology. To respect the values and the strategy of the luxury carmaker, **each brand should organize a Delphi study then finalize the results with a face-to-face** between the participants so that they can discuss the results, This scenario should be accompanied by the development of a clear road-map with the main features and the possibility of bespoke personalization.

Recommendation

The following road-map may be suggested on the development and commercialization of vocal in-car features.



Vocal in-car road-map proposition

Sources:

CHRISTENSEN Clayton (1997) The Innovator's dilemma - Harvard Business School Press.

Frost & Sullivan (2016) Women as Vehicle Customers North America.

Frost & Sullivan (2016 & 2018) Study Global Automotive Voice and Speech Recognition System Market.

MIDLER Christophe, MANIAK Rémi, BEAUME Romain (2012) Réenchanter l'industrie par l'innovation [Re-engaging industry through innovation (Trad.)] – Dunod.

Voicebot.ai - Voice UX Best practices (2018).

Voicebot.ai - In-car voice assistant consumer adoption (2020).

SémioConsult® is a consulting agency relying upon an internationally accepted expertise, mostly related to luxury business with its codes and specificities.

Specialised in brand positioning and identity management, as well as in anti-counterfeiting strategies, it includes in its client portfolio official entities along with prestigious French and Italian brands. www.semioconsult.com