How Mercedes is Creating Innovative Multi-Modal Experiences with Cerence Look

12月 7, 2021

By Tim Jungheim, Marketing Manager

Looking to the future of in-car assistants, we firmly believe that voice is more powerful when combined with the other senses: sight, touch, gesture and more. That's why we recently introduced Cerence Look, which enables drivers and passengers to interact with points of interest outside the car. For example, a driver can simply look at a coffee shop and ask, "What is the name of that restaurant?" or "What is that café on my left?" Imagine how transformative this is to the driving experience, extending the capabilities of the automotive assistant and furthering our goal of minimizing driver distraction.

Innovative automaker Mercedes-Benz and Cerence jointly worked on bringing this exciting technology to life: Mercedes' highly personalized and intuitive infotainment system, MBUX, uses the revolutionary Cerence Look technology in their Travel Knowledge feature, thereby leading innovation in multi-modal interaction. We recently sat down with Alexander Schmitt, Head of Speech Technology at Mercedes Benz AG, to hear a bit more about how he and his team are working to bring the best possible driving experience to Mercedes' customers.

Tim Jungheim: Thanks for taking this time to chat with us! Can you start by telling us a bit about your background and your role at Daimler?
Alexander Schmitt: Thank you for giving me the opportunity! I am a computer scientist holding a PhD in machine learning and spoken dialog systems. I have been working in speech dialog systems in both research and the corporate world for 15 years now. I still recall my first research project in 2006, where we started to teach mobile phones to understand spoken language based on a distributed speech recognition architecture. That was five years prior to the introduction of Apple's Siri. I am always thrilled seeing the rapid development of voice technology, and I am excited to be part of that journey.

Today at Mercedes-Benz AG, I lead a great team of Data Scientists, researchers, software developers, and project managers contributing to the next generation of the MBUX Voice Assistant. You may also know this system under the name "#HeyMercedes." One of my team's missions is the development of novel and innovative domains.

TJ: Could you please share a bit about how Travel Knowledge came to be and the motivation for bringing this type of technology into MBUX?
AS: Our aim is to offer our customers the best driving experience. We want to develop our voice assistant in such a way that it plays an integral role here. Therefore, for the HeyMercedes MBUX Voice Assistant, use cases that are strongly tied to driving and traveling are key to an attractive portfolio.

Travel Knowledge is an excellent example how we can enrich and augment this driving experience. With Travel Knowledge we allow our customers to get information about what they see while driving, e.g. you may ask about restaurants or shops, but also sights, such as the Brandenburg Gate, a church, or a castle up on a hill such as Neuschwanstein that you discover while driving on the autobahn.

TJ: How do you see the role of voice and natural multimodal interaction within Travel Knowledge, and overall within MBUX, both today and in the future?
AS: What actually makes a key difference in automotive voice assistants in comparison to others is its deep integration with the sensors and actuators of the vehicles.

This deeper integration into the vehicle's capabilities allows us to benefit from gaze detection or even gesture recognition. Did you know that in our current vehicles, the rear-view mirrors are adapted according to in which mirror you are looking? There is no need to actively select the left or right one anymore.

Making use of the information where the user is looking or pointing to is one of our key goals. This renders the interaction even more natural for Travel Knowledge, similarly as in this mirror example.

Check out this video for more from my conversation with Alex.

TJ: From a development and engineering perspective, what was your biggest learning during the development of Travel Knowledge?
AS: Data quality is key. If you want to get information about a shop next to you – say it's a drug store – and the system offers you information about a bakery because the data is not up to date, the experience is not satisfying. The same holds for opening hours. Where available, we also tell the users if the shop, restaurant or any other business they are asking for is currently open. The validity of this information heavily depends on if the owners are actively providing the correct information to our content providers.

TJ: Aside from Travel Knowledge, what are some other applications you could see for Cerence Look in the near term and beyond that could further enhance the in-car experience?
AS: There is still so many things that you discover while travelling that we do not have an answer for yet. How long is this tunnel, what is the name of this bridge, how tall is this building, what is the name of this river? These are just a few examples.

Moreover, we should target the ability to answer questions that are more specific. Instead of asking, "Tell me something about this bakery," our customers should be enabled to ask, for example, "Do they sell cakes or coffee?" Again, you see the critical loop to data availability and quality. However, that is definitively a journey we should strive for.

For more from my conversation with Alex, check out this [video](#), and learn more about Cerence Look [here](#).