THE DRAGON DRIVE INNOVATION SHOWCASE:

ADVANCING THE STATE-OF-THE-ART IN AUTOMOTIVE ASSISTANTS

Julien Villemure, Peter Yeh, Robert Policano, Michael Kaisser, Rainer Gruhn

Nuance Communications, Inc.
{Firstname.Lastname}@nuance.com

Abstract: The Dragon Drive Innovation Showcase (DD-ISC) is a demonstration platform used by Nuance Automotive to display the latest of its automotive assistant innovations. It feeds into Dragon Drive, Nuance's automotive platform that uses artificial intelligence to bring conversational personal assistants to over 200 million vehicles on the road today across more than 40 languages. DD-ISC combines a wide range of cutting-edge features under one hood, many of which are still in prototype stage and will typically reach market readiness in one to three years. A strong focus of DD-ISC is highlighting to the industry and broader public the contributions that new artificial intelligence and multi-modal technologies can bring to the advancement of the state of the art in intelligent personal assistants. Some of these technologies are: Personalization, which uses machine learning to model the preferences of individual users; Contextual Reasoning, which considers the current situation along with general world knowledge when making recommendations; Voice Biometrics, which identifies the speaker using voice "fingerprinting"; Always-On Listening, which removes the need for a push-to-talk button or a wake-up word; and Gaze Detection, which tracks the user's line of sight. In this paper we describe these technologies and how each plays its part in realizing a truly intelligent automotive assistant for drivers and passengers alike.

1 Introduction

Virtual assistants have been among the most visible technological trends in recent times. Amazon Alexa, Google Now, Apple Siri and Microsoft Cortana are well-known voice assistants offering to help users in many daily situations. Typically, such assistants are deployed on users' smartphones or in their houses, often as smart speakers. Yet there is one often overlooked environment where voice-first personal assistants can be especially powerful: the car.

Commonly, while driving, a user's hands need to be on the wheel, and their eyes need to be on the road; thus many forms of human-computer interactions (e.g. touch screens) are highly inconvenient or even dangerous in an automotive setting. However, as a driver's mouth and ears are less solicited in most driving situations, the use of voice input and output as a means of communication has a much smaller impact on cognitive load and contributes to a safer and less stressful driving experience. To this end, Nuance has been providing personal voice assistant technologies to car manufacturers, enabling them to create their own automotive assistants for more than a decade [1].

In this paper we describe the Dragon Drive Innovation Showcase (DD-ISC), a demonstration platform used by Nuance Automotive to display the latest of its automotive assistant innovations.

2 Background and Related Work

The assistant space in 2018 is dominated by two types of assistants: those found on mobile devices, such as smartphones, and those that primarily serve the home space [2], all of which are typically based on client-server architectures [3].

On mobile devices, Apple's Siri [4] is the most well-known assistant, present on all of Apple's mobile products, primarily iPhones and iPads. Siri was first released in 2011. Samsung followed in 2012 with its own assistant, S Voice [5], and later launched Bixby [6], another more intelligent take on a virtual assistant, in 2017. In the home, Amazon Alexa [7] is currently the leading solution, initially released in late 2014. Google followed in 2016 with a similar product, the Google Home [8]. Additionally, there are a lot of smaller players in the assistant market and many more products announced. Deutsche Telekom, for example, is set to launch its own Smart Assistant called Magenta [9].

In terms of functionality, the aforementioned assistants offer roughly comparable features. Typically, music can be controlled, appointments and alarms can be set up, the latest news or weather report can be requested. On top of utilities like these, Smart Home use cases motivate the second prominent set of functionalities, including turning the lights on or off, changing the temperature by setting a thermostat, etc.

Nuance Dragon Drive [1] differentiates itself from other solutions in that it allows our customers to build voice-controlled assistants that are automotive first: decidedly focused on automotive scenarios and uses cases. Accurate and personalized recommendations contribute to a reduced cognitive load for the driver, which translates to improved safety for all occupants. A hybrid architecture provides embedded functionality even without a network connection, and advanced cloud features when a connection is available, including e.g. real-time fuel prices and opening hours of businesses.

3 Architecture Overview

The DD-ISC platform combines voice, touch screen, and eye tracking to form a cohesive multi-modal interface. The voice component features automatic speech recognition (ASR), voice biometrics, natural language understanding (NLU), dialog management, and text-to-speech (TTS).

Dragon Drive's architecture is hybrid by design, combining embedded and cloud technology in the best possible way to maximize feature availability, latency, and ultimately user satisfaction. Typically, speech is processed by both the embedded and cloud recognizers in parallel, but only one module will return results to the user: voice commands that do not require cloud connectivity are processed locally by the embedded system, while queries that need to leverage the more powerful AI modules or that need access to real-time data—such as the weather forecast or fuel prices—are processed in the cloud. One central strength of such a hybrid approach is the ability to retain some functionality even when a network connection is unavailable. Another benefit is improved responsiveness for the use cases that do not require advanced cloud processing.

Figure 1 illustrates the overall DD-ISC architecture as a flowchart. Input to the system consists of user commands (voice or touch), learned user preferences (see *4 Personalization*), and contextual information about the car and the environment (see *5 Contextual Reasoning*). Based on this, a dialog manager determines the best action to take. If there is insufficient information to complete a task, a prompt is issued (on-screen and/or via TTS) in order to obtain further input from the user.

The cloud dialog manager uses the Nuance Reasoning Framework (NRF), which leverages several semantic technologies to reason about user preferences and contextual information, with the goal of determining any additional implicit constraints that should be considered when making recommendations [10]. As needed, NRF invokes an Intelligent Knowledge Access (IKA) module in order to retrieve relevant world knowledge and provide the best results. This may include reasoning about which content sources to use, applying the constraints inferred by NRF from preferences and context in addition to those explicitly mentioned by the user, and finally, relaxing the constraints if they cannot all be satisfied simultaneously.

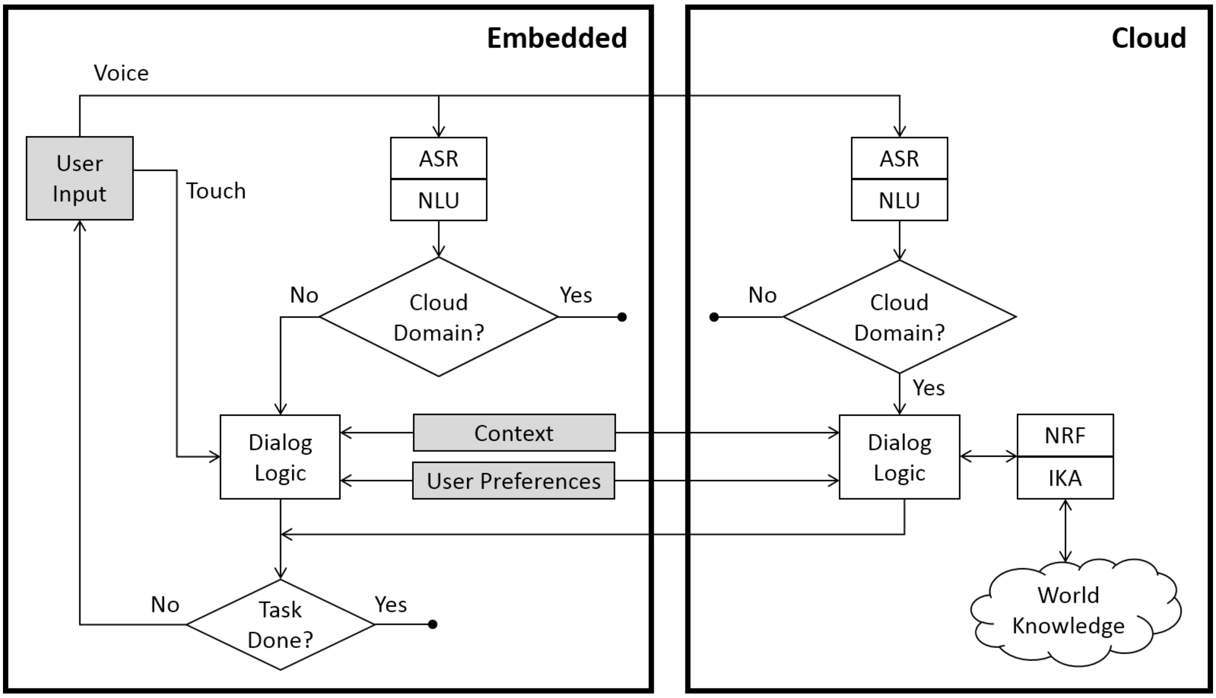

Figure 1 – Flowchart of DD-ISC illustrating multi-modal input within a hybrid embedded/cloud architecture. Output modalities (on-screen display and TTS) are omitted for simplicity.

4 Personalization

As mentioned previously, speech is a modality that is well suited for human-machine interaction in automotive settings. Yet, speech carries a few inherent challenges, one of them being the fact that it is a linear medium: the information presented to the user needs to be consumed by the user in that order, and the user has little to no influence on what information is presented.

Besides dialog, which allows users to ask for information in which they are specifically interested, personalization is an important mechanism through which issues stemming from this linearity problem can be solved. The knowledge that the system possesses about specific users can be used to present them only with those pieces that are known to strike their interest, or that are crucial for their decision making. This can help users to make better informed choices more quickly, and contributes to a better overall experience.

To give one practical example: when a specific user is looking for a restaurant, there are a multitude of properties that influence the choice he or she will make. On one hand, there are the properties of the restaurant itself, e.g. cuisine type, price range, driving distance, etc. On the other hand, there are the user's personal preferences. While *User A* might be a huge fan of spicy Chinese food, *User B* might not like spicy food at all. What personalization can bring to such a scenario is to learn the relevant user preferences over time, in order to not only make highly-targeted, personalized recommendations, but to also decide which details about the restaurant are presented to the user, so that he or she can make an informed decision for or against each recommended restaurant.

Preference learning is not limited to the user's favourite restaurant categories. Examples of other preferences that can be learned include price range, driving distance, and favourite brands, to name a few.

5 Contextual Reasoning

In addition to personal preferences and search criteria explicitly enounced by the user, the assistant relies on contextual information from the car's sensors and navigation system, along with general world knowledge from the cloud, in order to recommend the businesses most suited to the user's needs. Contextual cues leveraged by the system are described in this section.

5.1 Current Location

The geocoordinates of the car's current location are used to search for businesses in the vicinity of the user. For example, when the user says *"Drive to the grocery store"*, it is preferable to recommend grocery stores within a short driving distance, rather than several kilometres away.

5.2 Planned Route

When the navigation system is active and providing directions to the user in order to reach a destination, the assistant relies not only on the car's current location, but also on the full route plan, including the geocoordinates of the destination and a geographic plot of the route. The driving distance to the destination and the estimated time of arrival are also known.

Knowledge of the planned route enables the assistant to effectively increase the search area and recommend businesses that may be farther away, but nonetheless *on the way* to the destination, in an effort to minimize the incurred *detour distance*. For example, a business that is 50 km away but just 1 km off the planned route will result in a short detour, while a business that is only 10 km away but 5 km off the planned route will result in a longer detour.

In order to keep the search delay under an acceptable latency threshold, the search area normally does not span the entire route. Rather, the assistant uses world knowledge to intelligently select an adequate subsection of the route on which to restrict the search. Areas that are too far from cities and towns are excluded from search, as it is unlikely that we will find businesses "in the middle of nowhere". The assistant is however careful not to exclude the service areas typically found along highways, even in remote locations. This behaviour is illustrated by Figure 2:

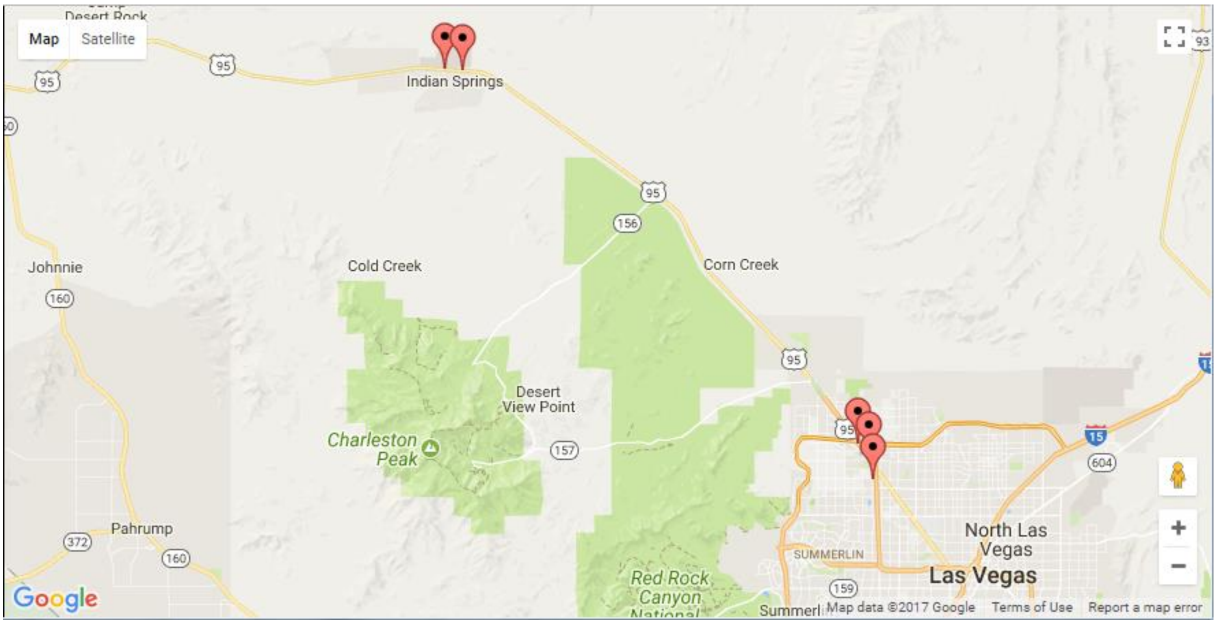

Figure 2 – The red pins indicate the search locations for filling stations along the route going from the Las Vegas Strip to Death Valley via Interstate 95 in Nevada, USA.

When navigating from the Las Vegas Strip to Death Valley National Park in Nevada, USA, the assistant searches for filling stations near the outskirts of the city and around the small town of Indian Springs where a service area is located. The neighbourhood around the car's current location in the heart of Las Vegas is excluded because it is not necessary to stop for refilling *right now*. No search is performed in remote desert areas along the highway because they are too far from any population centre or service area. By searching only in relevant locations, the assistant gives the user a choice between two reasonable options: filling up before entering a long stretch of highway, or stopping later at the next service area.

5.3 Temporal Context

When the opening hours of a business are known, it is possible to determine whether it will be open at the time we will arrive there. This lets the assistant exclude closed businesses from recommendations.

For example, if the user wants to find parking close to the destination, the assistant will only recommend parking lots that will be open at the time of arrival. This is particularly useful during early morning or late-night road trips, when most businesses would be closed at the present moment, but would open at a later point of the trip.

The *expected duration of stay* may be known as well. For example, the user may explicitly request parking for 2 hours. In the case of refueling at a gas stations, it can be assumed that the stay will be relatively short, maybe up to 15 minutes. This information is taken into account when determining whether a business will be open. A parking lot that closes 1 hour after arrival will not be recommended for a 2-hour stay.

5.4 Weather

The weather forecast can be accessed by the assistant to determine whether covered parking (e.g. a parking garage) is preferable, as opposed to on-street parking or open-air parking lots. Temperature and the risk of precipitation are the pieces of information leveraged in this case.

5.5 Car Profile

The characteristics of the car itself are important when recommending certain types of businesses. For instance, the height of the car is essential in determining whether it is possible to enter a certain parking garage.

Another example is fuel. Different cars run on different sources of energy: petrol/gasoline (sometimes mixed with varying quantities of ethanol), diesel, natural gas, electricity, etc. The assistant is aware of the energy sources supported by the car, and will only recommend filling stations that offer at least one of them.

In addition, the user can indicate a preference for "premium" or "super" fuel, commonly found at petrol stations alongside "regular" fuel. If such a preference is indicated, the assistant is able to better estimate the cost of filling up with the desired fuel type at a given station. Prices vary across fuel types and stations; it is useful to know which price to consider when ranking stations from cheapest to most expensive.

5.6 Car Sensors

Readings by various sensors mounted on the vehicle can trigger reminders and recommendations, as well influence the parameters of business searches.

For example, the assistant knows the vehicle's fuel autonomy, i.e. how far it can drive until it runs out of fuel or electricity. Based on this, the assistant can give the user a timely reminder to stop for refueling or recharging at an adequate station along the route. Typically, a search for filling or charging stations is centered around a very conservative estimate of the location where the car will begin to run out of fuel, i.e. well before the "out-of-fuel" indicator lights up. This accurately reflects the behaviour of most drivers who, in the vast majority of cases, prefer not to stop for refilling until necessary.

In the earlier Las Vegas route example (Fig. 2) the fuel autonomy is known to be sufficient to reach the town of Indian Springs. If that weren't the case, then the assistant would ignore the town and would instead search exclusively around Las Vegas and its outskirts.

Other examples include a low level of windshield wiper fluid, which can trigger a reminder to the user that stopping at a nearby filling station or convenience store is recommended, and active windscreen wipers, which indicate that precipitation is falling and therefore that covered parking options should be prioritized by the assistant when searching for nearby parking.

6 Voice Biometrics

By using voice "fingerprinting" technology, the automotive assistant is able to identify who is speaking in order to provide a more personal experience to drivers and passengers. Specific to a user's profile are not only car parameters such as seat configuration, but also location references like "home" or "work", and personal preferences which can be taken into account when searching for businesses (see *4. Personalization*). Biometric identification is also used as a security measure to protect access to personal data.

Multiple users can be signed in simultaneously, e.g. the driver and the passengers. The combination of voice biometrics and speaker localization enables the assistant to know exactly who is speaking and in which seat they are sitting. This is useful, for instance, if a user says *"I'm cold"* – the assistant can then adjust heating for the user's seat according to his or her climate control preferences.

User profile creation follows a simple enrollment process whereby the user types a name using the on-screen touch keyboard and repeats a chosen passphrase, e.g. "Hello Dragon", three times. A four-digit personal identification number (PIN) is also configured. Once enrollment is complete, the user may sign in to the system by speaking the passphrase or entering the PIN using the touch keyboard.

7 Always-On Listening / Just Talk

In the past, invoking the assistant required pressing a button or using a wake-up-word. This is no longer necessary with the "Just Talk" feature, which lets occupants engage more freely and intuitively with the assistant. Utterances not directed at the assistant are ignored. This results in a considerably more "humanized" feeling when interacting with the assistant.

8 Gaze Detection

As vehicle manufacturers continue to incorporate novel sensory technologies into their products, the resulting new inputs will need to be integrated with the automotive assistant to provide the most seamless, safe and enjoyable user experience. DD-ISC demonstrates how one of such technology, infrared-based eye tracking, can be tightly woven into the assistant's multi-modal architecture. The driver can simply glance at a building or other object situated within his or her field of view outside the vehicle, and query the assistant with questions and requests such as *"What is the user rating of that restaurant?"* — *"When does that business close today?"* — *"Call that store"* — etc.

9 Conclusion

We have described the Dragon Drive Innovation Showcase (DD-ISC) and positioned it within the context of the broader virtual assistant market, highlighting the principal ways in which Dragon Drive differentiates itself for the automotive realm. Our intuitive multi-modal interface contributes to a reduced cognitive load for the driver, resulting in a safer and more enjoyable driving experience. Powered by a hybrid embedded/cloud architecture, our assistant provides functionality that is both robust and responsive. Advanced contextual reasoning and personalization ensure that users receive recommendations tailored closely to their needs. Finally, with the help of voice biometrics and always-on-listening, an intuitive and seamless dialog with the assistant becomes possible.

References

[1] NUANCE COMMUNICATIONS, INC.
https://www.nuance.com/mobile/automotive/dragon-drive.html
Accessed January 2018.

[2] VERTO ANALYTICS INC.
http://www.vertoanalytics.com/verto-index-personal-assistant-apps
Accessed January 2018.

[3] GRUHN, R., K. TAKASHIMA, T. MATSUDA, A. NISHINO, and S. NAKAMURA: *CORBA-Based Speech-to-Speech Translation System*. ASRU, pp. 355-358, 2001.

[4] APPLE INC.
https://www.apple.com/ios/siri
Accessed January 2018.

[5] SAMSUNG ELECTRONICS CO., LTD.
http://www.samsung.com/global/galaxy/what-is/s-voice
Accessed January 2018.

[6] SAMSUNG ELECTRONICS CO., LTD.
http://www.samsung.com/global/galaxy/apps/bixby
Accessed January 2018.

[7] AMAZON.COM, INC.
https://www.amazon.com/Amazon-Echo-And-Alexa-Devices/b?node=9818047011
Accessed January 2018.

[8] GOOGLE LLC.
https://store.google.com/product/google_home
Accessed January 2018.

[9] DEUTSCHE TELEKOM AG.
https://www.telekom.com/en/media/media-information/consumer-products/with-smart-speaker-your-home-listens-to-your-command-508276
Accessed January 2018.

[10] JAIN, P., P. Z. YEH, E. STORY, J. VILLEMURE, D. MARTIN, and W. JARROLD: *Nuance Reasoning Framework*. ISWC 2017.