Voice interactions with multiple users

Let every passenger have a say
Table of contents

Introduction ........................................................................................................................... 3

ASR for all passengers ......................................................................................................... 4
Wake-up Word ...................................................................................................................... 4
Barge-in ............................................................................................................................... 4
JustTalk ............................................................................................................................... 4

Rising to the challenge with Nuance SSE ........................................................................... 5

Use cases for multi-user setups .......................................................................................... 8

Company background ........................................................................................................ 10
About Nuance ..................................................................................................................... 10
About Nuance Automotive ................................................................................................. 10
Introduction

Today's automotive systems feature rapidly progressing support of the driver. State-of-the-art automatic speech recognition (ASR) and text-to-speech (TTS) technologies provide for an easy and natural activation of the system and ensure highly reliable, conversational reactions to user commands at all times.

With Artificial Intelligence (AI) features, this is taken one step further: The automotive assistant is not only talking and reacting in a natural way, the reactions become more and more intelligent. Driver preferences are memorized and connected with other knowledge and without much interaction, the assistant comes up with a helpful suggestion – for the preferred parking place, restaurant, or playlist.

All this comes along so easily and naturally that communicative expectations for the assistant's potential are rising: A system that reacts so smoothly to the voice of the driver surely can handle the commands and questions of other passengers as well – or so is the assumption. Use cases for an extended performance of the automotive assistant are obvious: Passengers are interested in a POI along the route, they ask for specific music titles or they need to adjust the air conditioning.

In the following, this whitepaper will focus on various aspects of voice interactions with multiple users, such as activation techniques and the specific challenges of the multi-user mode in a car environment. The whitepaper also introduces Nuance's Speech Signal Enhancement (SSE) and its key benefits for multi-user setups. Finally, you will learn about potential use cases and an actual multi-user application for the car.
ASR for all passengers

To activate automatic speech recognition (ASR) in a multi-user setup, the same techniques are used as for just one user. The following sections introduce and explain the techniques briefly.

Wake-up Word
Just like for a single user, a so-called wake-up word can be defined for the multi-user setup – a feature that is already supported by Nuance technology. When the automotive assistant “hears” the wake-up word, it becomes active immediately and starts a speech session. It is then of no importance who speaks the word, driver or passenger.

Example:
The defined wake-up word is “Hello Dragon”. One of the passengers wants to know about the weather in Frankfurt. She simply asks:

“Hello Dragon, what’s the weather in Frankfurt?”

It is also possible to first wake up the system with “Hello Dragon” and then ask the question in a separate utterance after a brief pause.

Barge-in
The barge-in feature, also supported by the current Nuance technology, is closely connected to the wake-up word: The automotive assistant is alert even while speaking, so users can interrupt long prompts and thus shorten the interaction. As with the wake-up word, all passengers have access to this feature and can take over the lead in a voice interaction.

JustTalk
With the feature JustTalk, the activation process becomes even more natural: Instead of reacting to just one wake-up word, the system can be activated by several specific commands or even by all commands. This feature will become available in the course of 2018. It requires a tight integration of audio system and dialog setup and should therefore be deployed together with the Dragon Drive framework.
Rising to the challenge with Nuance SSE

If the assistant’s activation and use is just as easy for multiple users as it is for a single user, where is the challenge?

Possibly the greatest challenge is the noise/sound environment in the car. Specific car noises and travelling sounds mingle with the output of the infotainment system itself and the ongoing conversations between the passengers. In this wild mixture, the automotive assistant has to recognize the wake-up word and the subsequent command, ignoring all the other sounds in the car. While this is quite a task in a single-user setup, it is even more demanding with multiple users, where the commands can come from any direction and in any user’s voice and intonation. For these reasons, a high-quality speech signal is at the heart of any ASR automotive solution. This precondition can only be fulfilled if a number of speech signal processing technologies are used together.

Nuance Speech Signal Enhancement (SSE) has been highly successful over the last 15 years in combining state-of-the-art signal processing technologies that enable Dragon Drive to deliver high-quality speech recognition results. Find out in the following sections about the individual technologies and how they help dealing with the challenge of an automotive multi-user setup.
Elaborate microphone setups and configurable acoustical zones

The sound basis of the whole speech recognition process is a high-quality, robust microphone signal – no matter what seat in the car it comes from. Nuance SSE therefore supports a large number of dedicated microphone configurations, from single microphone setups for one or two passengers to multi-microphone setups that include all seats. The supported setups are highly configurable and cover a variable number of acoustical zones and up to 8 microphones.

Examples of supported microphone configurations

Speaker separation and multi-zone processing

Multi-user setups must also be flexible with respect to who can speech-control the system. There are many situations when only the driver should be in control, and other situations when only the back-seat passengers need to be excluded from voice interactions. Technologies like Beamforming or Passenger Interference Cancellation (PIC) methods help to separate individual speakers and ensure the flexibility in activating/deactivating acoustic zones. Like the microphone setup, acoustic zone processing is highly configurable and can deal with numerous speech situations in the car.

Examples of acoustic zone configurations with Beamforming / PIC
For more information on microphone setups, Beamforming and PIC configurations please contact your Nuance representative.

**State-of-the-art signal processing technologies**
Signal processing technologies remove noise from the microphone input and send out a cleaner signal for improved speech recognition. In this process, noise effects caused, for example, by fans, open windows or road bumps, are eliminated.

**Multi-channel Acoustic Echo Cancelling (AEC)**
With this technology, signals coming from the infotainment system will be removed. This makes sure that the wake-up word really wakes up the assistant.

**Closely integrated speech enhancement and ASR system**
The integration ensures a robust system performance in any circumstance.
Use cases for multi-user setups

An excellent speech signal and the ensuing stability of the speech recognition system will first and foremost improve the user experience, be it for a single user or for multiple users in a multi-user setup. In addition, the high signal quality paves the way for new use cases that involve more than one speaker. Consider some examples:

**Interactive entertainment solutions**
Surviving long journeys by car is not only difficult with children on board. Interactive speech-controlled game applications help to fight boredom, to have fun as a group, and even to include the driver if the traffic circumstances allow for it. With Dragon Tunes, Nuance has developed a speech-based music game for a multi-user setup. The game, that follows the rules of “Name That Tune”, was presented during CES 2017 in a six-seat minivan.

![Dragon Tunes, Step 1: Selecting avatars](image)

After an avatar is assigned to each player the game starts with music playback. At the same time, the speech recognition is active and listens for the wake-up word “Got it”. The passenger who calls out “Got it” can now say the name of song
and artist. During that utterance, the acoustic zone only for this passenger will be active. After the successful naming of
song/artist, the game will resume with a playback of a new tune.

**Telephone calls involving all passengers**
Such calls include conference calls in a business context but also private calls between family at home and travelling
family members.

**Access to the speech-driven infotainment system**
All passengers can be given access to the car's infotainment system with its many features. And just as easily, the access
can be restricted to the driver if circumstances ask for it.
Company background

About Nuance

Nuance Communications, Inc. (NASDAQ: NUAN) is seen as the leading provider of voice and language solutions for businesses and consumers around the world. Its technologies, applications and services make the user experience more compelling by transforming the way people interact with information. Every day, millions of users and thousands of businesses experience Nuance’s proven applications and professional services.

Nuance is reinventing the relationship between people and technology through speech and language solutions driven by advances in Artificial Intelligence and cognitive computing. It has pioneered the evolution of speech recognition technology that today integrates Artificial Intelligence (AI) to transform the way people interact with the devices, systems, apps, and services that surround them. Every day, millions of people and thousands of organizations experience our technology through intelligent systems that can listen, understand, learn, reason and facilitate life and work. Our clients span large companies and organizations, including hospitals, banks, airlines, carriers and car manufacturers that leverage our technologies and services to make businesses and products run more smoothly and create a better experience.

Speech is one of the most natural and intuitive ways to interact with devices, applications and systems, lessening our reliance on the mouse, keyboard and touchscreen. We have developed a broad portfolio of speech recognition and Natural Language Understanding (NLU) technologies that integrate machine learning and big knowledge for the variety of systems and services that leverage virtual and collaborative assistant offerings across devices and services in the Mobile, Enterprise and Healthcare industries. Further, our Document Imaging business drives increased productivity and security for the world’s largest enterprises that need to gain control over document capture and workflows.

About Nuance Automotive

Speech recognition, NLU, AI and predictive touch solutions from Nuance have pioneered many of the personal assistant technologies and intelligent systems in the devices we use every day from the world’s leading brands – including mobile devices, cars, televisions, wearable devices, and now the emerging ecosystem of the Internet of Things. We deliver a more human experience with technology, keeping consumers better connected and informed – consistently adapting to and predicting their needs.

The Nuance Automotive business delivers automotive-grade solutions enabling drivers all over the world access to information and services and providing the safest, smartest and most natural user experience. Nuance’s voice technology has been shipped in more than 160 million cars from Ford, Toyota, BMW, Mercedes, Fiat and other major automakers and is at the heart of over 14 million connected car experiences on the market today. Nuance’s Dragon Drive provides the industry’s most comprehensive suite of solutions for the connected car, giving automakers and suppliers the ability to integrate a natural language voice interface, content, and connectivity that is customized for their individual brand.