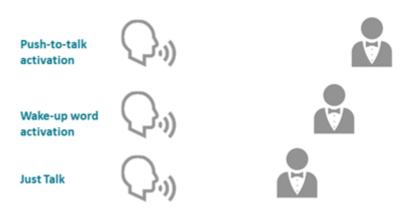## Hey, Is Anybody Listening?

September 22, 2021

By Dr. Holger Quast, Product Strategy, Core Products

If you've ever used voice recognition in your car, chances are you've pressed that little button with the talking face icon on your steering wheel to activate the voice assistant. In more advanced systems, you may have uttered the increasingly familiar wake-up word, which thanks to their frequent usage on smartphones and smart speakers, are becoming commonplace in automotive as well. You get the assistant's attention with a phrase like "Hey (name your favorite brand here)," then speak your question or command. It's hands-free, no buttons needed, and thus makes the interaction more natural.

I like to think of the usage of the push-to-talk button vs. wake-up word activation as a metaphor for calling a butler. Imagine it's 1762; you are the Earl of Sandwich and fancy a snack. You pull a heavy brocade string to ring a bell. Your butler enters, asking, "how may I help you, sir?" and you say, "bring me a glass of champagne / my loafers / your finest cut of meat between two slices of toast." The string-activated bell is a bit like the push-to-talk button: it would be more convenient to always have your butler in the room with you, so you can simply request, "James, please bring me my snack," akin to activating the voice assistant with its wake-up word.

Taking this analogy one step further, imagine speaking a series of requests in a short sequence: "James, bring me one of these snacks that you make." "James, I'd like a bit of salad, too." "James, put it here." "James, no, here on the table." "James, I want to have some wine with that." "James…James!" This shows us that while a wake-up word has many advantages over the push-to-talk activation, there are some situations in which speaking the wake-up word every single time is quite awkward. James doesn't need to be reminded each time you speak that you are speaking to him – if you and James are the only people in the room and you say, "can you make me one of those lovely meat-between-bread bites again?" James will perfectly understand he was being addressed, even though his name was not spoken.

Moving some 250 years ahead, the learning from the scenario above also holds true for automotive user interfaces. If you are alone in the car and say, "turn on the radio," it's clear you addressed the voice assistant. To whom else would you be talking?

Push-to-talk activation

Ring for personal assistant
Dedicated effort in different modality to start voice assistant

Wake-up word activation

Call personal assistant
Dedicated effort in same modality (voice) to start assistant

Just Talk

Personal assistant is always available
Waiting in background for command, no activation needed.

The butler analogy: different degrees of proximity of your assistant

Intrigued by this scenario – after all, we love our sandwiches, and we love our voice assistants – our Cerence DRIVE Lab team set out to properly understand drivers' desired behavior by conducting a series of experiments. We tested how people prefer to start the interaction with an assistant and made some interesting observations: if people are in the car with other people, they tend to address each other by name – "hey Vanessa, can you pass me a sandwich?" "John, quit making such a ruckus!" Similarly, they usually address the voice assistant by name, activating it by wake-up word: "Hey [assistant name], play some music."

In some cases though, it's clear by the command you are saying that a request is intended for the voice system – for example, "Set the navigation screen to north up mode" – so a wake-up word would not necessarily be needed, even in the multi-occupant case. Further, and not surprisingly, we found that when the driver is alone in the car, the most natural way to phrase a command is without having to speak the wake-up word first.

To do justice to these preferences, we have built Cerence Just Talk. You can speak a command, with or without the wake-up word, and no need to press any buttons. Cerence Just Talk analyzes what you are saying and whether it forms a command or question, and only if that's the case, it jumps into action. Otherwise, it remains quiet in the background – just like James would.

Next to the analysis Cerence Just Talk performs on the textual content level, we apply machine learning to statistics such as intonation features of the voice, the number of passengers in the vehicle, and more, to make sure the system understands the situational context and can make smart decisions about when to speak up.

Play Video

CERENCE

JUST TALK

While we may be poking some fun at the standard "Hey X" wake word in the video above, of course, just like push-to-talk buttons, they continue to have their raison d'être in speech dialogues depending on driving situation and user preferences. We actually have a number of nifty tricks for wake-up word activation: with our technology, you can customize your in-car assistant by giving it a name you like – something like "Schatzi," for example – so you can further personalize your experience. We can even set up the system so the wake-up word can be recognized in different positions in the utterance, such as, "play some music for me Schatzi" or "I'm hungry, Schatzi, please find me a restaurant".

And if you have multiple assistants available in your car, should you ever get confused and call the assistant by the wrong name – for instance, speaking a home control command (intended for the assistant you use on your smart speaker at home) but using the wake-up word of your car's assistant – she won't get upset but instead will automatically route your request to her virtual colleague.

Cerence Just Talk, however, marks the next step for user interfaces. No button press or wake-up word needed. Your voice assistant knows when you are talking to it, making the interaction even easier and more human-like and creating an intuitive, natural, helpful, and enjoyable user experience.

To learn more about Cerence Just Talk, visit https://www.cerence.com/cerence-products/core-technologies.