Why Industrial Voice Crushes Siri, Google & Alexa

by DAN PAUL, Customer Success Lead, January 15, 2018

Speech recognition technology is now a part of daily life. With U.S. revenue growth projections of 16% annually, and a recent RetailDIVE piece reporting that fully 40% of consumers are expected to use a voice assistant by 2021, expect to find yourself talking to more and more machines over the course of your day. We have all had experience with automated phone systems and/or the myriad of personal electronic devices with speech as the main user input (Alexa, Siri, Google, etc.). Unfortunately, the user experience with these consumer-grade options leaves much to be desired. "I didn't quite get that" messages and confounding translations ("10 percent" = "temper sent"??) can leave users ready to launch their devices out the window.

Anyone using the latest industrial-grade voice recognition technologies will tell you that there exists a marked difference between consumer-grade speech recognition software and today's industrial-grade offerings. We have identified the top four reasons for this improved performance.

1. Accuracy
    a. As you might imagine, speech recognition is challenging. While great strides are being made in natural language processing (NLP), consumer systems still rely on something known as 'Speaker Independent" recognition. This means that the users of the system never train the system on how they talk and moreover, most of these systems do not improve over time as they speak more into it. With the wide array of accents, grammar, and speech patterns it takes massive computing power and a massive data set of examples to determine what someone is saying - and the mistakes are many with this variability.
    b. High-end Industrial Voice Systems use 'Speaker Dependent' recognition, which means that each individual user trains the system with their own voice. This may

sound like a large time investment, but most voice systems need only 50-70 phrases for typical operational workflows and training takes less than 30 minutes. User-specific training means that a person's unique vocal patterns, accents, and language are captured for use determining the attempted command or response.

2. Hardware/Microphone
    a. Consumer-grade speech recognition typically relies on picking up sound from a distance or from over a phone. In both cases capturing good quality audio is challenging because of the varying volume levels, likelihood and presence of background noise, and varying quality levels of the device itself.
    b. The best industrial speech systems rely on rugged headsets to bring the microphone close to the source of the desired sound: the mouth. In addition, microphone arrays cancel out background noise before it even reaches the speech processor. Normalizing these components of sound (volume/gain, quality) is fundamental.
3. Performance
    a. Most consumer-grade speech systems require connectivity to the internet to function. Assuming a strong and continuous connection, the response time for these systems can be pretty good due to the supercomputers crunching this sound data in the background. But any break in connectivity brings your efforts to a halt.
    b. In contrast, many industrial systems perform speech recognition right on the device worn by the user. This means that you are able to use the system even on a desert island. Also, the system is designed to allow a limited array of phrases at any given time, ensuring that the recognition of these phrases is instantaneous- every time.
4. Adaptive Recognition
    a. As previously stated, the data being collected and compiled for consumer-grade recognition is enormous and growing. This helps raise overall recognition scores but does little for the individual experience.
    b. In contrast, industrial systems can adapt to changing speech patterns. As the day wears on, voices change somewhat. Hay fever or a common cold can drastically alter speech. By taking constant samples of recognition scores and adjusting the underlying speech template for users, the system improves over time- even with changes in the user's speech.

Speaker dependent voice-directed workflows in industrial settings, now in use for over two decades, have become the de-facto standard in DC technology. Mountain Leverage has been delivering voice solutions for over a decade to a fiercely loyal customer base who enjoy improvements in accuracy, productivity, training time, safety, employee satisfaction, and more. We understand how speech recognition works and use it as a tool to deliver amazing results across many industries.

Dan Paul is VP of Customer Success at Mountain Leverage. With over a dozen years of experience delivering voice solutions, he has a passion for helping companies discover and unlock their operational excellence. Reach him at voice@mountainleverage.com.