

What’s next



Enterprise

Over and Out – Moving beyond the walkie-talkie voice interface: Part II of “What’s left to tackle in voice technology”

When game changing technology comes out sometimes it's hard to look past the dazzle and see what else is needed to make it great. This is true for today's smart speakers and voice interfaces in general. At Nuance, we are focused on transforming the way people interact with technology – looking beyond what's cool to what makes tech interfaces “easy” to use. This post will examine the shortcomings of today's smart speakers and the complexity of getting to a viable solution.

Ken Arakelian

Posted January 3, 2018



“Over” was short for “over to you” indicating that it’s your turn to talk on a short wave radio or walkie-talkie (or any half duplex comm tech for you nerds out there). Smart speakers are super cool and a step forward in voice – but they’re still half duplex, klunky, unnatural voice interfaces. We’ll all look back one day and remember how quaint today’s smart speakers were – like we remember morse code, tape players, and VCRs. Try turn-taking a face-to-face conversation or conference call sometime, and you’ll get a feel for what smart speakers, and all voice interfaces for that matter, are missing out on. There’s a whole field of study around the protocols and rules of human conversation called “Pragmatics” that study how humans interact one to one, one to many and many-to-many.

For example – I’ll say, “Alexa, play ‘Fool in the Rain’ by Led Zeppelin on Spotify,” and wait the requisite three seconds of silence so Alexa knows I’m done talking (might be easier to just say “over”). Then Alexa says, “I’m sorry, I can’t find ‘Fool in the Rain’ by Led Zeppelin on Spotify.” I’ll remember I cancelled Spotify and try to correct myself by speaking over Alexa, “No, play it on Amazon Music.” It’s natural to do this – a person wouldn’t miss a beat having the same conversation.

In addition to the half duplex limitation, Alexa also can’t understand multiple speakers. Even the best user interfaces today employ turn-taking to manage the conversations and don’t work at all with more than two speakers in a conversation. For example, if my children interrupt Alexa while she’s playing “Fool in the Rain” and ask her to play “[Space Unicorn](#)“, a song that

can make you insane after hearing it for the 400th time, I typically respond by shouting, “Laa Laa Laa Laa!!” to confuse Alexa and keep her playing good music.

Managing the turn-taking in a conversation with multiple speakers is no simple task. It requires that you listen while you talk and also respond to visual queues (in a face-to-face conversation). For example, Japanese speakers often produce back-channel expressions such as *un* or *s* while their partner is speaking. They also tend to mark the end of their own utterances with [sentence-final particles](#) and produce vertical head movements near the end of their partner’s utterances. See [Turn-taking – Wikipedia](#) for a long description of the complexity. The listen-and-talk problem gets exponentially worse when you add more speakers. A bot will need to know whether it’s having a friendly conversation and should wait until the person is done talking, or if the bot is arguing and should cut into the rant.

Recognizing these short-comings is the first step in over-coming them. Nuance R&D is working on these problems and others to transform the way people interface with technology.

Stay tuned for Parts 3 and 4 as we catalog the technical problems when telling your customer to “talk to the IVR like you would a human”.

- [Part 1 – Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo: Part 1 of What’s left to tackle in voice technology](#)
- [Part 3 – Oh! I know! I know! – Proactive voice interfaces](#)

Tags:

More Information

	<p style="text-align: center;">Unleash the power of voice technology</p> <p style="text-align: center;">Nuance’s market-leading voice technologies enable your organization to create interactive voice response (IVR) and speech solutions so innovative, they’ll get your customers talking.</p> <p style="text-align: center;">Learn more</p>
---	---



About Ken Arakelian

Ken Arakelian is the director of product management for Nuance's On Demand business. With more than 15 years of experience, Ken has worked in the contact center industry as a consultant, account manager and product manager.

[View all posts by Ken Arakelian](#)