

Voice control is taking off, but it's not taking over yet

January 24 2017, by Fraser Allison



Credit: AI-generated image ([disclaimer](#))

If unexpected packages start showing up at your door, you might want to have a word with one of your smart devices.

Earlier this month, a six-year-old in Dallas asked her family's [Amazon Echo](#) smart speaker for a dollhouse. And Alexa, Amazon's Siri-like

artificial assistant, promptly [ordered one to their house](#).

A San Diego TV news show picked up the story, and inadvertently repeated it when one of the news anchors commented: "I love the little girl, saying 'Alexa ordered me a dollhouse'." Overhearing this, several other Amazon devices in homes across San Diego [attempted to buy more dollhouses](#).

The story might sound ruefully familiar to anyone who has tried to have a conversation with Apple's Siri or Microsoft's Cortana. Our devices have become quite good at listening to us, but that doesn't always mean they understand.

Researchers at Microsoft recently pinpointed this as a potential problem with today's talking interfaces: they are marketed as "intelligent" assistants, with clever jokes and worldly knowledge, yet they often frustrate us with their lack of common sense.

In a [small study](#), the researchers found that the people who continued to talk to their digital assistants over time were those who had started out with the lowest expectations.

What does a voice interface actually do?

When you speak to a voice interface, it has to:

- "hear" the sound of your voice, and distinguish it from background noise
- figure out where each word begins and ends, ignoring your "umms" and "ahhs" match the sound of each word to a word in the dictionary, picking the right one from context if there are [homophones](#)
- correctly interpret the meaning of the whole sentence

- generate a meaningful and useful response that matches your request.

Each one of these is a complex technical challenge, and different technology companies have made progress in different areas.

Google Now is good at giving relevant responses to a wide range of requests because it benefits from Google's troves of data about the web, and your personal activities, if you use Google services.

Amazon Echo is particularly good at hearing your requests from across a noisy room, thanks to a noise-cancelling far-field microphone array. Of course, it's also good at making purchases through Amazon.

Over the past few years, voice interfaces have become much better at understanding everyday or "natural" speech rather than only stilted and carefully worded commands. They are still better at handling simple queries, like "who's playing in the Australian Open?", and tend to struggle with more complicated requests, like "who's playing in the Australian Open for the first time this year?", and follow-up questions, like "will it rain during the finals?".

The situation is even more mixed for languages other than English: while Siri supports more than 40 languages and dialects, so far Alexa is only available in English and German. But all of these features are steadily improving.

Where voice interfaces stutter

So will voice interfaces soon take over all of our technology, as predicted in the film [Her](#)? Gartner, a technology research firm, [has forecast](#) that by next year, 30% of our interactions with technology will be conversations with voice-enabled interfaces.

But voice interfaces have limitations, and not all of them can be solved by better technology.

Noise pollution is one major hurdle. Can your device distinguish what you're saying from the background noise around you? Technology can help with that, including noise reduction, personalised voice recognition and lip reading.

But what about the background noise you're creating for others by talking to your smart device? Imagine a person sitting next to you at the office – or on an aeroplane – chatting to Siri while you're trying to read, and you can see why voice interfaces may not always be socially acceptable.

Another set of issues come from the mental demands of voice interfaces. Learning to use a voice-based system can be hard, especially if there is no screen, as with Amazon Echo.

If you've ever called up a bank or a telephone company, you know the miserable combination of concentration and boredom that comes from listening to a synthesised voice list out all your options while you wait for the one you need and try not to mix them up. Traditional graphical interfaces avoid this problem by showing you the available options and letting you quickly tap your choice.

After you've learned voice commands, using them can be distracting. Researchers have found that voice commands [derail your train of thought](#) more than a mouse and keyboard.

This is particularly dangerous for in-car voice interfaces: a pair of studies from the University of Utah found that drivers were [distracted for up to 27 seconds](#) after using [voice commands](#).

Finding its voice?

So voice interfaces are unlikely to take over entirely, but they will find useful niches in our lives. They are already common in cars, where they will hopefully become less distracting as the technology improves.

In the kitchen, you can ask Alexa to talk you through a recipe or update your shopping list while your hands are busy cooking. In virtual and augmented reality, voice interfaces can let you control the system when you can't see your hands at all.

In language learning, they can be used for practicing pronunciation. Most importantly, voice interfaces help users with motor impairments, RSI or dyslexia to overcome their disabilities.

Voice interfaces are a long-awaited technology, and there are good reasons to think their time has finally come. Just remember that they may not yet be as clever as they sound. And you might want to put a PIN code on voice purchases if children are around.

This article was originally published on [The Conversation](#). Read the [original article](#).

Provided by The Conversation

Citation: Voice control is taking off, but it's not taking over yet (2017, January 24) retrieved 2 May 2024 from <https://phys.org/news/2017-01-voice.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.