Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants

Matthew B. Hoy

Published online: 12 Jan 2018.

Submit your article to this journal

View related articles

View Crossmark data

Routledge
Taylor & Francis Group

Check for updates

EMERGING TECHNOLOGIES

Matthew B. Hoy and Ariel F. Pomputius, Column Editors

Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants

Matthew B. Hoy

Mayo Clinic Libraries, Mayo Clinic, Rochester, Minnesota, USA

ABSTRACT

Voice assistants are software agents that can interpret human speech and respond via synthesized voices. Apple's Siri, Amazon's Alexa, Microsoft's Cortana, and Google's Assistant are the most popular voice assistants and are embedded in smartphones or dedicated home speakers. Users can ask their assistants questions, control home automation devices and media playback via voice, and manage other basic tasks such as email, to-do lists, and calendars with verbal commands. This column will explore the basic workings and common features of today's voice assistants. It will also discuss some of the privacy and security issues inherent to voice assistants and some potential future uses for these devices. As voice assistants become more widely used, librarians will want to be familiar with their operation and perhaps consider them as a means to deliver library services and materials.

Introduction

People have wanted to talk to computers almost from the moment the first computer was invented. Science fiction is full of computers that can hold a conversation, from HAL 9000 and the Starship Enterprise's computer to Marvin the Paranoid Android and KITT the car. Just a few decades ago, the idea of holding meaningful conversation with a computer seemed futuristic, but the technology to make voice interfaces useful and widely available is already here. Several consumer-level products developed in the last few years have brought inexpensive voice assistants into everyday use, and more features and platforms are being added all the time. Users can do everything from asking simple informational questions to playing music and dialing their phone or turning lights on and off via voice control. This column will explore the basic workings and common features of today's voice assistants. It will also discuss some of the privacy and security issues inherent with voice assistants and potential future uses for these devices. As voice assistants are more widely used, librarians will want to become familiar with the

technology, which has the potential to be a delivery mechanism for library materials and services.

What Are Voice Assistants?

Simply put, voice assistants are the realization of the science fiction dream of interacting with our computers by talking to them. Apple's Siri, Microsoft's Cortana, Amazon's Alexa, and Google's Assistant are all software agents that run on purpose-built speaker devices or smartphones. The software constantly listens for a key word to wake it up. Once it hears that key word, it records the user's voice and sends it to specialized server, which processes and interprets it as a command. Depending on the command, the server will supply the voice assistant with appropriate information to be read back to the user, play the media requested by the user, or complete tasks with various connected services and devices. The number of services that support voice commands is growing rapidly, and Internet-of-Things device manufacturers are also building voice control into their products.

Apple's Siri assistant has been around the longest, released as a standalone app in 2010 and bundled into iOS in 2011. Microsoft followed shortly thereafter with Cortana in 2013. Amazon launched Alexa with its Echo-connected home speaker in 2014, and Google's Assistant was announced in 2016 along with its Home speaker and is also embedded in the Google app for Android-based smartphones. Each assistant has its own unique features, but the core functions are the same. Voice assistants differ from earlier voice-activated technologies in that they can respond to a much larger number of commands and questions. This is because they are always connected to the Internet; each interaction is sent back to a central computing system that analyzes the user's voice commands and provides the assistant with the proper response. Earlier voice-activated devices relied on a smaller set of "built-in" commands and responses. Recent advances in natural language processing, also known as computational linguistics, has allowed voice assistants to create meaningful responses quickly. Hirschberg and Manning credit these recent improvements in natural language processing to four things:

> (i) a vast increase in computing power, (ii) the availability of very large amounts of linguistic data, (iii) the development of highly successful machine learning (ML) methods, and (iv) a much richer understanding of the structure of human language and its deployment in social contexts.[1]

As personal computers have grown cheaper and more powerful, and people have created more and more online text to be analyzed, scientists have used that text to train voice assistants to listen and respond to our requests in more natural and meaningful ways. Voice assistants can parse requests phrased in a number of different ways and interpret what the user is most likely to want.

For example, to ask Google's Assistant to remember where one parked his or her car, a user can say any of a number of phrases: "Remember where I parked," "I parked here," "I left the car on 6th street," or "the car is in the south lot" will all get a similar result. Google will remember where the user parked the car and, when asked later, will be able to respond accordingly. The user can ask questions in a similarly natural way; asking "where did I park," "where did I leave the car," or "do you remember where I parked" all trigger the expected response. Natural language processing avoids the user frustration of earlier voice recognition systems, which required specific phrases and patterns in order to work properly.

What Can Voice Assistants Do?

Although each currently available voice assistant has unique features, they share some similarities and are able to perform the following basic tasks:

- send and read text messages, make phone calls, and send and read email messages;
- answer basic informational queries ("What time is it? What's the weather forecast? How many ounces are in a cup?");
- set timers, alarms, and calendar entries;
- set reminders, make lists, and do basic math calculations;
- control media playback from connected services such as Amazon, Google Play, iTunes, Pandora, Netflix, and Spotify;
- control Internet-of-Things-enabled devices such as thermostats, lights, alarms, and locks; and
- tell jokes and stories.

In addition to these tasks, voice assistants can add other features, often called "skills," that expand their abilities by interfacing with other programs via voice commands. Amazon's Alexa has skills for playing Jeopardy, ordering your usual drink from your local Starbucks, and summoning an Uber or Lyft using connected account data. Google's Assistant has similar skills but lags behind Amazon in the sheer number of available skills, largely due to being released later.[2] These skills are built by third-party developers, similar to the way apps are developed for smartphones. Google Assistant also integrates with several tools that allow users to create their own skills. Using web services like Tasker and IFTTT (If This Then That), users can craft skills that will allow them to automate social media posts, turn devices on and off, and hundreds of other possibilities. For example, telling Assistant "Good morning" could launch a number actions designed to speed up the user's morning routine: turning on the coffee maker, reading the news and calendar events, opening the garage door and starting the car, and then locking the doors, arming the alarm, and adjusting the thermostat after occupants have left the house for the day. All of the devices necessary to perform these tasks

remotely are available today; a voice assistant simply provides the bridge that allows users to issue commands verbally rather than via an app.

Voice assistants are available on a wide variety of hardware platforms. Amazon and Google both market dedicated home speaker devices for their voice assistants. Amazon makes several variations of its Echo product, from the tiny Echo Dot to the Echo Show, which has audio and video capabilities. Google's speaker hardware is called the Home and also comes in mini- and full-size models. Apple is just entering the home speaker market, with the announcement of its Siri-enabled HomePod device, scheduled to be available in December 2017. Microsoft has focused on building Cortana into Windows 10 PCs and phones and recently partnered with Harmon Kardon to develop a Cortana enabled home speaker.

Assistants are available on most smartphone platforms as well; Google's assistant is integrated into Android phones and can be installed as a separate app on the iPhone, although some features are disabled.[3] Amazon's Alexa has Android and iOS app versions, and Microsoft and Amazon are working together to bring Cortana to Amazon devices and Alexa to PCs.[4] Apple has kept their assistant off of non-iOS devices, but Siri is available on all Apple devices, including iPhones, Macbooks, iPads, and the Apple Watch. As the voice assistant market stabilizes, it seems likely that there will be additional integration and that feature sets across the main voice assistants will become similar. For the moment, Amazon is the dominant player in the field, due to launching a home product first with a large media library available out of the box. Google is building capacity, and the addition of a home-based speaker and integration into other Google products will drive their market share up. Apple may also become more of a contender with the release of HomePod late in 2017 and the addition of more Apple-branded connected home products. Microsoft is not likely to gain much traction, as their share of the smartphone market is negligible and they lack a compelling home-based product.[5]

Security and Privacy

While voice assistants have interesting and useful features, they also pose several unique problems. One of the main issues with these voice-activated devices is security. Anyone with access to a voice-activated device can ask it questions, gather information about the accounts and services associated with the device, and ask it to perform tasks. This poses a major security risk because these devices will read out calendar contents, emails, and other highly personal information. In one reported case, a man discovered that the iPad in his living room would unlock the front door for anyone who stood outside and asked Siri to let them in.[6] Google has recently upgraded its Assistant software to include voice printing, which uniquely identifies each user by voice

and prevents the device from reading out personal information. Apple is also teaching Siri to recognize a user's voice, but that feature had not yet been released at the time of this writing. Amazon's Alexa is just as prone to these security issues, and Amazon is working to deploy a similar voice printing system. Alexa has the added issue of being built into Amazon's store interface. By default, anyone with voice access to the device can order items using the owner's Amazon account. There are options to set a voice passcode to confirm purchases, and all goods will ship to the owner's address on file, but there is still potential for malicious users to purchase goods on the owner's account. Household members could make unauthorized purchases as well, like the six-year-old who ordered herself a dollhouse and four pounds of sugar cookies via Alexa.[7]

Voice assistants are also vulnerable to several other attacks. Researchers have recently proven that voice assistants will respond to inaudible commands delivered at ultrasonic frequencies.[8] This would allow an attacker to approach a victim, play the ultrasonic command, and the victim's device would respond. There is also the possibility this type of attack could be embedded in broadcast media. A product advertisement on television that contains embedded ultrasound commands to add the item to your shopping cart might spur you to buy it. The idea seems far-fetched, but a series of news stories about the previously mentioned girl that ordered herself a dollhouse caused issues for many people. The newscaster spoke the words, "Alexa, order me a dollhouse" in the course of reporting the story, and hundreds of Alexas in viewer homes responded.

Privacy is another major concern for voice assistant users. By their very nature, these devices must be listening at all times so that they can respond to users. Amazon, Apple, Google, and Microsoft all insist that their devices are not recording unless users speak the command to wake the assistant, but there has been at least one case where a malfunctioning device was recording at all times and sending those recordings back to Google's servers.[9] Even if the companies developing these voice assistants are being careful and scrupulous, there is a potential for data to be stolen, leaked, or used to incriminate people. A murder investigation in Arkansas lead to the authorities issuing a warrant to Amazon to retrieve the suspect's Alexa history.[10] Children's privacy when using these devices is also a concern; Mattel recently shelved plans to build a voice assistant aimed at children, code named "Aristotle," after complaints were raised that the device would be invasive of children's privacy.[11]

Potential Future Uses

Voice assistants have the potential to radically change how users interact with computers. For many users, the ability to read and type is a barrier to

accessing information. Voice assistants can bridge the information gap for those users. Recent research has shown that voice assistants can benefit dementia sufferers by providing an ever-present voice that can answer the same questions again and again without losing patience and offer encouragement when needed.[12] For others, reading the instructions their physician provides can be difficult. Building these abilities into currently available consumer technologies would be much more cost effective than a purpose-built device, and many users would already be comfortable operating these devices. Voice assistants could also read books and other long-form documents to users. Although they still sound somewhat robotic, the vocal qualities of voice assistants are rapidly improving. Once they improve enough to not be off-putting, every book could be an audiobook.

Voice assistants also have the potential to revolutionize translation. Google recently announced a new set of earbuds that pair with its voice assistant for real-time voice translation.[13] Users launch the application by asking the assistant to help them speak a language. The user's phone captures audio spoken by the other party, relays it to Google's translation servers, and plays the translated version in the user's earpiece. When the user wants their speech translated, he or she presses a button on the earpiece, which sends the audio to the server, where it is translated and relayed to the listener via the speaker on the user's handset. While Google's translation is notoriously error-prone, particularly for medical or colloquial language, the near-instant results should make it useful for simple conversations.[14]

Voice assistants may be useful for library promotion and management as well. There are already tools available that allow libraries to create skills for voice assistants that list events at the library in local community calendars. Coding additional features that would allow patrons to hear hours and announcements, renew their items, hear what new items are available, or schedule a consultation with a librarian would be relatively simple. More complex tasks, such as searching databases and requesting interlibrary loans, are probably out of the scope of a tool that can only provide feedback via voice.

Voice assistants could also easily be programmed to act as virtual tour guides in smaller gallery or exhibit spaces. Patrons could ask the assistant to tell them about an exhibit, and the assistant can read back prepared remarks. Libraries with a technology focus may want to consider lending these devices and providing basic training so that patrons can experiment with these devices in their homes as well.

Conclusion

The complexity and accuracy of voice recognition technology and voice assistant software have grown exponentially in the last few years. Currently available voice assistant products from Apple, Amazon, Google, and

Microsoft allow users to ask questions and issue commands to computers in natural language. There are many possible future uses of this technology, from home automation to translation to companionship and support for the elderly. However, there are also several problems with the currently available voice assistant products. Privacy and security controls will need to be improved before voice assistants should be used for anything that requires confidentiality. Librarians should monitor these products and be ready to provide assistance to their patrons with these devices. They should also explore the possibilities for providing library materials via voice assistants as the technology matures.

Notes on Contributor

Matthew B. Hoy, MLIS, AHIP (hoy.matt@mayo.edu) is Associate Director of Libraries, Mayo Clinic, 200 1st Street SW, Rochester, MN 55905.

References

1. Hirschberg, Julia, and Christopher D. Manning. "Advances in Natural Language Processing." *Science* 349 (6245): 261–266. doi:10.1126/science.aaa8685.
2. Moore, Clayton. "The Most Useful Skills for Google Home." *Digital Trends* (May 3, 2017). https://www.digitaltrends.com/home/google-home-most-useful-skills/.
3. Lee, Nicole. "Google Assistant on the iPhone Is Better than Siri, but Not Much." *Engadget* (May 17, 2017). https://www.engadget.com/2017/05/17/google-assistant-iphone-hands-on/.
4. Hachman, Mark. "The Microsoft-Amazon Deal Leaves Cortana Speakers with One Advantage: Skype." *PCWorld* (August 30, 2017). https://www.pcworld.com/article/3221284/windows/the-microsoft-amazon-deal-leaves-cortana-speakers-with-one-advantage-skype.html.
5. Spence, Ewan. "Windows Phone Is Dead, Long Live Microsoft's Smartphone Dream." *Forbes* (July 12, 2017). https://www.forbes.com/sites/ewanspence/2017/07/12/microsoft-windows-phone-windows10-mobile-strategy/.
6. Tilley, Aaron. "How A Few Words to Apple's Siri Unlocked a Man's Front Door." *Forbes* (September 21, 2016). https://www.forbes.com/sites/aarontilley/2016/09/21/apple-homekit-siri-security/.
7. Liptak, Andrew. "Amazon's Alexa Started Ordering People Dollhouses after Hearing Its Name on TV." *The Verge* (January 7, 2017). https://www.theverge.com/2017/1/7/14200210/amazon-alexa-tech-news-anchor-order-dollhouse.
8. Zhang, Guoming, Chen Yan, Xiaoyu Ji, Tianchen Zhang, Taimin Zhang, and Wenyuan Xu. "DolphinAttack: Inaudible Voice Commands." In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, New York, NY, 103–117. doi:10.1145/3133956.3134052.
9. Tung, Liam. "Google Home Mini Flaw Left Smart Speaker Recording Everything." *ZDNet* (October 11, 2017). http://www.zdnet.com/article/google-home-mini-flaw-left-smart-speaker-recording-everything/.
10. Buhr, Sarah. "An Amazon Echo May Be the Key to Solving a Murder Case." *TechCrunch* (December 27, 2016). http://social.techcrunch.com/2016/12/27/an-amazon-echo-may-be-the-key-to-solving-a-murder-case/.

11. Vincent, James. "Mattel Cancels AI Babysitter after Privacy Complaints." *The Verge* (October 5, 2017). https://www.theverge.com/2017/10/5/16430822/mattel-aristotle-ai-child-monitor-canceled.

12. Wolters, Maria Klara, Fiona Kelly, and Jonathan Kilgour. "Designing a Spoken Dialogue Interface to an Intelligent Cognitive Assistant for People with Dementia." *Health Informatics Journal* 22 no. 4 (December 1, 2016): 854–866. doi:10.1177/1460458215593329.

13. Palladino, Valentina. "Google Pixel Buds Are Wireless Earbuds That Translate Conversations in Real Time." *Ars Technica* (October 4, 2017). https://arstechnica.com/gadgets/2017/10/google-pixel-buds-are-wireless-earbuds-that-translate-conversations-in-real-time/.

14. Patil, Sumant, and Patrick Davies. "Use of Google Translate in Medical Communication: Evaluation of Accuracy." *BMJ* 349 (December 15, 2014): g7392. doi:10.1136/bmj.g7392.