Alexa vs. Siri vs. Cortana vs. Google Assistant: A Comparison of Speech-based Natural User Interfaces

Gustavo López, Luis Quesada, Luis A. Guerrero

University of Costa Rica, San José, Costa Rica
{gustavo.lopez_h, luis.quesada}@ucr.ac.cr, luis.guerrero@ecci.ucr.ac.cr

Abstract. Natural User Interfaces (NUI) are supposed to be used by humans in a very logic way. However, the run to deploy Speech-based NUIs by the industry has had a large impact on the naturality of such interfaces. This paper presents a usability test of the most prestigious and internationally used Speech-based NUI (i.e., Alexa, Siri, Cortana and Google's). A comparison of the services that each one provides was also performed considering: access to music services, agenda, news, weather, To-Do lists and maps or directions, among others. The test was design by two Human Computer Interaction experts and executed by eight persons. Results show that even though there are many services available, there is a lot to do to improve the usability of these systems. Specially focused on separating the traditional use of computers (based on applications that require parameters to function) and to get closer to real NUIs.

Keywords: Human Factors · Intelligent Personal Assistant · Smart Device

1 Introduction

Alexa, Siri, Cortana and Google Assistant are the most common and extensively used Intelligent Personal Assistants available in the market. From the academic perspective, they are Speech-based Natural User Interfaces (NUI).

In the Human-Computer Interaction (HCI) domain, speech-based NUIs are systems that users operate through intuitive actions related to natural human behavior via voice instructions. There are many names for speech-based NUIs. However, this paper focuses on voice-activated intelligent personal assistants deployed in smartphones or smart speakers.

Nowadays, most smart-device manufacturers have their own voice assistant application. However, the designs vary significantly and their functionalities are very different [1].

The hypothesis of this paper is that the run of manufacturers to release new versions of their voice-activated personal assistants has had a large impact on their products, especially in their simplicity and usability.

An example of this assertion is that when voice-activated personal assistants were first released their behavior was limited and the commands were very rigid and structured [1]. Afterwards, several efforts were conducted to create more natural interfaces [2–5]. However, some of the current devices went back to rigid and structured commands.

This paper presents a functional and usability test of some of the most prestigious voice-activated personal assistants in the market (i.e., Amazon Alexa, Apple Siri, Microsoft Cortana and Google Assistant).

Each personal assistant answered the same request and a group of eight people rated each response on two categories: natural feeling and correctness. Results show that Google Assistant is the most natural personal assistant. However, it is also the less correct. On the other hand, Siri is the most correct personal assistant and the least natural.

The rest of this paper is arranged as follows: Sect. 2 illustrates the methods used in this work. In Sect. 3 the evaluation process is explained. Section 4 presents the results, which are further discussed in Sect. 5. Finally, Sect. 6 is the conclusion.

2 Evaluated Systems

In this paper, we describe an evaluation of four intelligent personal assistants. This section introduces the four evaluated systems: Amazon Alexa, Apple Siri, Microsoft Cortana and Google Assistant. These are the most popular assistants.

2.1 Amazon Alexa

Alexa is an intelligent personal assistant developed by Amazon. Alexa is linked with Amazon's smart speaker "Echo". According to Amazon, Echo is a hands-free, voice-controlled device that plays music, controls smart home devices, provides information, reads the news, and sets alarms, among other functionalities [6].

Moreover, Alexa is capable of voice interaction, music playback, making to do lists, setting alarms, streaming podcasts, playing audiobooks, and providing weather, traffic, and other real time information. Alexa can also control several smart devices using itself as a home automation hub [7].

Some of the promotional commands for Alexa include: "Alexa, find me a Chinese restaurant.", "Alexa, re-order paper towels". "Alexa, what's on my calendar today?", "Alexa, set a timer for 20 minutes", "Alexa, play Adele from Prime music", and "Alexa, what's my commute?" [6].

Additionally, Alexa services can be invoked from compatible devices developed by Amazon (i.e. Kindle Fire, Amazon Fire Phone and Fire TV) [7].

2.2 Google Assistant

Google Assistant is an intelligent personal assistant developed by Google. It was designed to allow conversational usage [8]. Google Assistant's predecessor is Google Now. Google Assistant is linked with Google Home, a voice-activated speaker [9].

Google Now uses a natural language user interface to answer questions, make recommendations, and perform actions by delegating requests to a set of services. Moreover, it delivers information to users predicting their requirements.

Some of the promotional commands for Google Assistant include: "Ok Google, Remind me to pick up a birthday card", "Ok Google, Book me a table for 6 at Quartino for 8:30", "Ok Google, Who invented sushi?" [8].

2.3 Microsoft Cortana

Cortana is an intelligent personal assistant created by Microsoft for Windows powered devices (both PC and mobile). Cortana can set reminders, recognize natural voice without the requirement for keyboard input, and answer [10].

Some of the promotional commands for Cortana include: “What’s the weather like?”, “Call Sarah”, “What is five miles in kilometers”, “What's the definition of 'table'?” [10].

2.4 Apple Siri

Siri is a personal assistant for Apple devices. Originally released in 2010, it is the oldest of the most used intelligent personal assistants. The most common uses for Siri are web browsing and dictation [11]. Currently, Siri is currently under a redesign process to add new functionalities and adapt it to new devices.

Some of the promotional commands for Siri include: “Show my photos from Utah last August”, “What movies are playing today?”, “Find videos I took at Iva’s birthday party”, “Text Pete ‘See you soon smiley exclamation point’”, “Read my latest email”, “Find a table for four tonight in Chicago” [11].

2.5 Summary

This section summarizes the devices and intelligent personal assistants evaluated in this paper. Table 1 shows the two evaluated smart speakers.

Table 1. Device summary

Device	Manufacturer	Release date	Current price	Connectivity
Echo	Amazon	Nov 2014	$179.99	Data, Bluetooth, Wi-Fi
Home	Google	Nov 2016	$129	Wi-Fi

Table 2. Personal Assistant Summary

Personal Assistant	Manufacturer	Release date	Supported Languages
Siri	Apple	2011	20
Cortana	Microsoft	2014	8
Alexa	Amazon	2014	2
Google Assistant	Google	2016	4

Google Home functions with Google Assistant as artificial intelligence, Echo works with Alexa. However, this paper also considers two intelligent personal assistants that do not have a specific device (i.e., Cortana and Siri). Table 2 summarizes the four evaluated intelligent personal assistants.

3 Methods

This study included the participation of eight subjects (average age = 26, StaDev ± 5, males = 6, females = 2) familiar with technology and the use of Intelligent Personal Assistants.

During the execution of the evaluation, all participants gather in the same place. They listened to the responses of each personal assistant. Table 3 describes the devices used for each personal assistant and Figure 1 shows the devices.

One of the main researchers followed a script of requests and repeated each request to each device separately in random order. Neither Text to Speech nor computer-generated speech were used during this research.

Table 3. Devices used to access each Personal Assistant

Personal Assistant	Device
Siri	IPhone 7 plus
Cortana	Microsoft Surface Book
Alexa	Echo Dot
Google Assistant	Google Home

Fig. 1. Devices linked to the intelligent personal assistants.

There are many classifications of assistant functionalities. To define the evaluation procedure we will use the following classification [12]: shopping and buying assistant, care assistant, travel and entertainment assistant and administrative assistant.

The script used in this research had four categories (i.e., shopping and buying assistant, travel and entertainment assistant, administrative assistant, and miscellaneous). The care assistant category only focused on functionality it was evaluated separately.

One of the researchers presented the devices with a request using the same voice tone and pace. All the participants listened to the answers and gave a response using a 5-point Likert scale for naturality of the response and correctness.

In the shopping assistant category, four features were considered: managing shopping lists, ordering products from the internet, finding restaurants or stores and finding store schedules.

For the travel and entertainment category, five requests were included: sports update, movie theater information, transportation, distance and arrival time providing two locations, and identifying a song.

In the administrative assistant category, six features were considered: managing multiple timers and alarms, to-do lists, reminders for the same date and a specific date, meeting scheduling and composing emails.

Finally, for the miscellaneous category 21 features were evaluated: newscasts, traffic updates, weather forecast, device management, games, random functions (flip a coin, random numbers, roll a die), conversions, measurements, currency, tip calculator, solving math problems, and other functionalities such as (telling jokes, beatboxing, answering random questions or telling facts).

To evaluate framework compatibility, a review was executed and the devices were tested with the smart home devices available for the research team.

4 Results

This section presents the results of evaluation described above. Also, compatibilities of the intelligent personal assistants are presented.

4.1 IFTTT Framework Compatibility

If This Then That (IFTTT) is a public platform of services that has gained much industry momentum and adoption. IFTTT provides a logical paradigm for controlling services. Users log onto the IFTTT platform and configure their available devices to be activated by a large set of triggers.

In this research, a web review was conducted to assess the compatibility of each intelligent personal assistant with IFTTT framework and the only personal assistant that does not have full compatibility with the framework is Siri. However, Apple is working on an update that will supposedly add this feature.

4.2 Smart Home Framework Compatibility

This research also reviewed the compatibility of the personal assistants with smart home devices (the available devices during the research). In this case, Alexa, Google Assistant and Siri were able to easily configure and access smart home devices. Finally, Cortana could not be configured. Reviewing Microsoft's website, this functionality is going to be added soon.

4.3 Evaluation Results

This section presents the main results of this research. The evaluation outcomes will be presented divided by features; afterwards, the results of each category will be

described. Finally, we will present the general results of this research. Table 4 shows the best and the worst personal assistant in both correctness and naturality for each one of the features evaluated.

Table 4. Best and worst personal assistants by feature. Abbreviations: S = Siri, C = Cortana, A = Alexa, GA = Google Assistant, N/A = Not applicable.

Feature	Best		Worst	
	Correctness	**Naturality**	**Correctness**	**Naturality**
Shopping and Buying Assistant				
Introductions	S	GA	C	A
Managing shopping lists	C	A	GA	GA
Shopping online	A	A	S, C, GA	S, C, GA
Finding restaurants	C, GA	C	A	S
Store schedules	C	C	S	GA
Travel and Entertainment Assistant				
Sport updates	A, S	S, C	GA	GA
Movie info	C	S, GA	A	C
Transportation info	S	A	GA	C
Transportation time	GA	GA	A	A, S
Identifying a song	C	C	GA	GA
Administrative Assistant				
Alarms and timers	A, S, GA	A, S	C	C
Managing to-do lists	A	A	GA	GA
Reminders	S, C	S	GA	GA
Reminders in a date	C	C	GA	A
Schedule a meeting	C	C	GA	GA
Compose an email	S	S	A	A
Miscellaneous				
Local news	C, GA	GA	S	S
Traffic updates	C	C	GA	GA
Weather forecast	A, S, C, GA	S	N/A	C
Device management	A	A	GA	GA
Games	GA	GA	S	S
Random (coin)	GA	GA	C	C
Random (dice)	A, S, C, GA	GA	N/A	S
Random questions	GA	GA	S	S
Random facts	S	GA	C	A
Conversions	GA	C, GA	A, S, C	S
Measurements	C, GA	S, GA	A	A
Currency	A, S, GA	GA	C	C
Holiday information	GA	GA	C	C
Calculate tips	GA	GA	A	A
Math problems	S, C, GA	GA	A	A
Time zones	GA	GA	C	C
Jokes	GA	C	S	S
Beatbox	GA	GA	A	A
Translation	GA	GA	A	A
Word definitions	A	A	S, C, GA	S, C, GA

In the shopping and buying category, the best personal assistant was Cortana. However, the only one capable of shopping online was Alexa. Alexa and Cortana gave the most natural responses in this category. The least correct personal assistants in this category were Siri and Google assistant; also, Google Assistant gave the least natural responses.

In general, for the shopping and buying category, the Echo dot, using Alexa was the most liked device. This was expected as it is designed for shopping. The request with the best response in all devices was to find a restaurant. Ordering things was only available on Alexa. However, the request with the most natural response was the store schedules. The request with the worst response was shopping online as only Alexa was capable of doing it.

In the travel and entertainment category, Siri and Cortana share the first place in correctness and Siri was the one with the most natural responses. Again, Google assistant was the least correct in this category. Cortana and Google Assistant gave the least natural responses. Also, in this category, the request with the best (correct) and most natural response in general was sports update.

In the travel and entertainment category, Google Assistant gave the best response to the location and arrival time query. Alexa gave the worst response when prompted to get information about a movie.

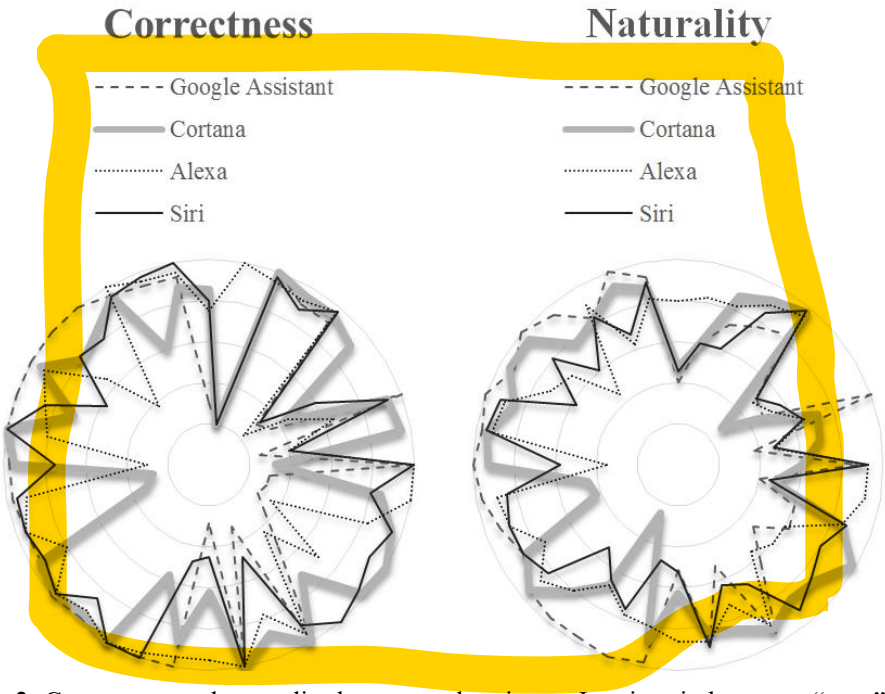

Fig. 2. Correctness and naturality by personal assistant. Interior circle means "poor" correctness/naturality; exterior circle means "excellent" correctness/naturality. Lines represent all tested features.

The administrative assistant category has a tie in the most correct personal assistant between Siri and Cortana. The most natural personal assistant was again Siri. Google Assistant was the least correct and natural personal assistant in this category. The best-responded query was timers and alarms in both correctness and naturality. The best response in general was the handling of timers. However, Cortana gave excellent results sending emails and setting reminders. Google Assistant had the worst response when prompted to set a reminder.

Finally, Google Assistant overwhelmingly won the miscellaneous category both in correctness and in naturality. The difference in this category is of large; the second place (Siri) has less than half the points that Google Assistant has for correctness. As for the naturality, Google Assistant won this feature by over 400%. Cortana gave the worst results in correctness for this category, closely followed by Alexa and Siri. Alexa and Siri also were the least natural.

In this category, the best-responded request was the conversions. Google Assistant had the best-rated response in 11 of the 21 categories (perfect score). Cortana had the worst response when prompted to know which day of the week a holiday was. Figure 2 shows the results.

The results presented in this section were unified giving one point to each personal assistant when they had the best or worst response.

In general, performing the statistical analysis no significant differences were found between the systems. In other words, the evaluation shows no preference for any system, neither in naturality nor in correctness. This result was extracted of an analysis of all the results together.

Siri was the most correct device; however, Google assistant was the one with the most natural responses. Google assistant suffered because it does not support some of the features evaluated in this research. Siri had the worst results in naturality.

5 Discussion

Science fiction has shown the concept of intelligent assistants through movies, TV shows and books. New technologies and algorithms have allowed the development of personal assistants for commercial use for some years. Available assistants provide a vast variety of functionalities. However, this variety depends on the personal assistant implementation and its purpose.

There is an opportunity for improvement in all tested devices. No trend was observed. All devices showed at least one functionality with wrong or unnatural answers. Nevertheless, Google Assistant was the best in miscellaneous category and the worst in the other categories.

A remarkable feature of Google Assistant was the naturalness of answering some questions. The tone and pace of the female voice used by the Google device expressed surprise, suspense and joy. These features were no always offered by Siri, Cortana and Alexa. Even though, some answers were catalogue as natural using these devices.

Siri and Cortana enhance theirs answers with visual information. Also, they show what phrase or words are recognizing. Google Assistant and Alexa can show additional information too, but using their respectively apps. This additional information was helpful in some features using maps, photos, or graphics.

Finally, new studies must be conducted. The intelligent personal assistants' potential could be tested in unexplored areas like counseling, marketing, learning, and sales. Also, combination of these devices and technologies with robots, data centers, and machine learning techniques provides new opportunities.

Acknowledgments. This work was partially supported by *Centro de Investigaciones en Tecnologías de la Información y Comunicación* (CITIC), *Escuela de Ciencias de la Computación e Informática* (ECCI) both at *Universidad de Costa Rica*. Grant No. 834-B6-178.

References

1. Ghosh, S., Pherwani, J.: Designing of a Natural Voice Assistants for Mobile Through User Centered Design Approach. In: Kurosu, M. (ed.) Human-Computer Interaction: Design and Evaluation. pp. 320–331. Springer International Publishing, Cham (2015).
2. Cowie, R., Cornelius, R.R.: Describing the emotional states that are expressed in speech. Speech Commun. 40, 5–32 (2003).
3. Dryer, D.C.: Getting personal with computers: How to design personalities for agents. Appl. Artif. Intell. 13, 273–295 (1999).
4. Becker, C., Kopp, S., Wachsmuth, I.: Why emotions should be integrated into conversational agents. In: Nishida, T. (ed.) Conversational Informatics: an Engineering Approach. pp. 49–68. John Wiley & Sons (2007).
5. Ball, G., Breese, J.: Embodied Conversational Agents. Presented at the (2000).
6. Amazon Inc.: Amazon Echo, www.amazon.com/echo%0A.
7. Amazon Inc.: Alexa Skills Kit, https://developer.amazon.com/public/solutions/alexa/alexa-skills-kit.
8. Google: Google Assistant, https://assistant.google.com/.
9. Google: Google Home, https://madeby.google.com/home/.
10. Microsoft Coorporation: Cortana, https://www.microsoft.com/en-us/mobile/experiences/cortana/.
11. Apple: Siri, http://www.apple.com/ios/siri/.
12. Sathi, A.: Cognitive Devices as Human Assistants. In: Cognitive (Internet of) Things: Collaboration to Optimize Action. pp. 29–39. Palgrave Macmillan US, New York (2016).