ATINER CONFERENCE PRESENTATION SERIES No: COM2018-0120

ATINER's Conference Paper Proceedings Series
COM2018-0120
Athens, 24 October 2018

Interacting with Intelligent Assistants to Predict Consumer Satisfaction

Rashmi G. Elera and D. C. Grant

Athens Institute for Education and Research
8 Valaoritou Street, Kolonaki, 10683 Athens, Greece

ATINER's Conference Paper Proceedings Series
COM2018-0120
Athens, 24 October 2018
ISSN: 2529-167X

Rashmi G. Elera, Undergraduate Student, University of Washington, USA
D. C. Grant, Lecturer, University of Washington, USA

Interacting with Intelligent Assistants to Predict Consumer Satisfaction

ABSTRACT

With the rise in popularity of intelligent assistants, there is an increasing need to understand and evaluate both strengths and shortcomings of the technology, in order to define specific areas for improvement and to understand where these interfaces are ideally suited. We describe the current state of personal digital assistants and evaluate performance by testing voice activated queries in four distinct categories including Translation, Current/Real Time Events, "How to" questions and General Knowledge. Experiments show that Microsoft's Cortana beat the two competitors with an impeccable accuracy of 100% followed by Amazon's Alexa with an average accuracy of 74% and Apple's Siri with only 49.8% accuracy. Siri was fastest to respond on the few questions it correctly answered, with an average speed of 2.09 seconds followed by Cortana with an average speed of 2.35 seconds and Alexa at the average speed of 2.63 seconds. Cortana had the highest accuracy and overall effectiveness. Analysis of these three assistants illustrates the current ability of intelligent assistants to aid consumers. This work also demonstrates tremendous potential of voice activated interfaces in the future. Evaluating which category each assistant performed best (or worst) can be a strong predictor of user satisfaction; essential for the future development of effective intelligent assistants. This research also reinforced concerns about a relatively poor ability of some voice-activated assistants to interpret the accents of non-native English speakers.

Keywords: human computer interaction, intelligent assistant, user satisfaction, user study, voice-activated interfaces.

Introduction

The rise of intelligent assistants comes from the ever-growing demand for faster and more efficient user and information systems. Making information more efficient and accessible is a major challenge HCI researchers face. This paper evaluates the existing state of human voice interaction with the analysis of three different voice activated personal assistants (Apple's Siri, Microsoft's Cortana and Amazon's Alexa).

These devices are able to access the Internet to provide general knowledge, entertainment, and answer many different types of questions. However, it is important to note that evaluating these assistants can be a challenge due to the diversity of the tasks they can perform which include texting other users, scheduling appointments and finding answers to general knowledge questions. Considerable time and resources are required to evaluate each of these tasks separately using several different metrics.

To overcome these difficulties, we developed an empirical approach, which strives to standardize the assessment of three distinct assistants and evaluate them in a just manner. We calculate product usability as well as consumer feedback to measure overall user satisfaction. Jiang et al (2015) conducted a study trying to "find and use correlations between explicit ratings of user experience and implicit behavioral signs by the users such as click and dwell time" [8, 9]. Although our experiment does attempt to find an association between user satisfaction and the effectiveness of digital assistants, implicit behavioral signs are not measured. We instead consider the effectiveness of these assistants as well as their capability for accuracy and rapid response.

Definition of 'Intelligent Assistants'

There are many different definitions of what an intelligent assistant truly is. Sharma defines intelligent assistants as "a software agent that can perform tasks or services for an individual" [10]. While Boros defines these assistants as "voice enabled human computer interfaces (HCI) that integrate automatic speech recognition, text-to-speech synthesis and natural language understanding" [11].

Several interchangeable terms describe intelligent assistants: "smart automated assistant", "voice activated assistant", "voice-controlled virtual assistant" and "smart mobile assistant", etc. Though there are several interfaces that use intelligent assistants, we limit our scope to intelligent assistants on home and mobile devices. We will also restrict this research to disregard queries that involve functions such as calling a person or logging into personal applications as these actions are not possible for all the intelligent assistants that we evaluated [15, 19].

The Intelligent Assistants

The devices used to perform these queries were Siri (iOS 10, iPhone 7), Cortana (version 2.4.6), Amazon (Echo Show version 594447320). If the task needed to access any device or applications (e.g. Translation), we completed installations prior to the experiment to make sure the tester would not encounter systemic problems that would influence the outcome of speed and accuracy results. All of the experiments took place in a quiet conference room, in order to reduce the disruption of environmental noises. Even though real-world environments often involve more background noise, competing voices and interruptions, we purposefully eliminated those factors to simplify the experiment and standardize results.

a) Siri: short for Speech Interpretation and Recognition Interface, is an intelligent assistant developed by Apple iOS. It uses a natural language user interface and voice queries to answer user questions. Siri also has the ability to make recommendations and perform actions within the iPhone. This intelligent assistant can execute several commands made by the users. Examples include sending messages, setting reminders, making a dinner reservation and many more. Siri is made of three components which include service delegation, personal context awareness and a conversational interface. Service delegation refers to the access to all of the built-in apps within the iPhone. Personal context awareness and conversational interface rely on Siri's ability to understand what the user is saying, word for word. Siri's intelligence and capabilities have definitely expanded since it first appeared in 2007.

b) Alexa: developed by Amazon, is also an intelligent personal assistant. It operates from Amazon's smart speaker called Echo and other devices. It provides users with several features including the ability to control smart home products, search the internet, report the weather, and much more. Alexa is a scalable cloud service that provides skills that help users interact with devices using voice in a natural way. It connects to the Internet and controls third party devices. Voice recognition as well as natural language processing are the artificial intelligence behind Alexa.

c) Cortana: the personal assistant developed by Microsoft uses the Bing search engine as well as data that is stored in the user's phone to make personalized recommendations. Cortana utilizes a user's data (email, calendar, and browsing history) to attempt and anticipate the user's needs. Much like Siri and Alexa, Cortana is also able to set reminders, provide answers to queries and more. However, although it uses speech recognition and understanding, Cortana also uses data mining and data gathering to provide answers and make suggestions.

Literature Review

Evaluating the Importance of Voice in Human Computer Interaction (HCI)

Since natural language communication clearly provides effortless and effective communication in human-to-human interaction, the immense potential for voice-directed interfaces in human-computer interaction should be obvious. Although natural language processing technologies have existed and evolved for over sixty years, most of the major technology breakthroughs in this field are recent. As computers have become more affordable and accessible, this accessibility has driven the increased importance of user interfaces that are both user friendly and highly efficient.

Intelligent assistant technologies are based on a combination of conversational intelligence and natural language processing. They are by definition "engineered entities residing in software that interfaces with humans in a human way" [13]. These assistants commonly integrate the functions of an array of applications in order to allow the consumer to use a single application interface, instead of separately interfacing with several different technology applications. A rise in artificial intelligence has driven these intelligent interfaces to replace common applications and act as a knowledge navigator. Natural language processing allows the system to interact with the users in a natural manner. This human to computer interaction can now simulate a natural conversation.

Recent HCI advancements now allow consumer's spoken communication with computer interfaces to be possible. This can be especially helpful when an interface user suffers from a visual impairment or experiences a physical difficulty that does not allow them to use the mouse or keyboard. Many types of physically disabled computer users should not have to depend on mouse, keyboard and screen, but should have a different range of options to access information.

An intelligent digital assistant coined "Project Nethra" was designed to provide visually disabled computer users a friendly and effective way to interact with technology [23]. Project Nethra provided access to social media and Internet services that have been largely inaccessible to the visually impaired using previous screen readers and similar applications. Despite its important improvements over other technologies, Nethra lacks the sophistication and third-party integration functionalities that mainstream intelligent assistants can now provide.

Intelligent assistants can now handle a wide variety of tasks, from making calls and managing a calendar, to answering general knowledge and current event questions, and even searching the Internet for information. By evaluating the usability and effectiveness of intelligent assistants, we can determine areas for improvement and extrapolate user satisfaction. In 2015 Jiang's work on "*Automatic Online Evaluation of Intelligent Assistants*" [8, 9] placed the varied requests into six categories for evaluation:

- Device control: launch applications and play music.
- Communication: make phone calls and send text messages.
- Location: find or navigate to certain places.

- Calendar: check calendar, create reminder and set alarm.
- Weather: check weather conditions.
- Other: all other supported requests.

That study used five representative products, "Apple Siri, Google Now, Microsoft Cortana, Samsung S Voice, and Nuance Dragon" to categorize requests, but only deeply analyzed usage for Microsoft Cortana. This deep analysis data strictly focused on use during beta testing by developers. We take a different approach to usability testing and analyze comparative capabilities for three consumer devices while used by consumers. Our approach more closely mimics 'real-world' use and provides realistic comparison between the devices which measures user experience and satisfaction.

Measuring and Evaluating User Satisfaction

Properly evaluating which intelligent assistant is most successful in providing correct answers and timely information sheds light on which interface has the most potential for user satisfaction. User satisfaction, as defined by Kelly in the 2009 work titled, "*Methods for evaluating interactive information retrieval systems with users*" [10], "can be understood as the fulfillment of a specified desire or goal". A more recent study conducted by Kiseleva [11] which collected participant responses based on their overall satisfaction with the task and the estimated effort it took, concluded "effort is a key component of user satisfaction across the different intelligent assistant scenarios". In the case of verbal interactions, less effort required to receive an answer for the queries almost always amounts to higher user satisfaction. Accuracy and the speed of response are the two most important performance factors/measures in determining whether users are truly satisfied with these intelligent interfaces.

Usability and consumer feedback on features are other significant factors used to measure overall user satisfaction. While in 1993, Dumas and Redish defined usability as "people who use the product can do so quickly and easily to accomplish their task" [5]. The International Standards Organization defined usability in 1994 as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" [12]. In 1998, Chinn developed a questionnaire for user interface satisfaction where participants rated familiar software products. He believed that subjective satisfaction of a system is a critical measure of the system's true success [4].

We calculate usability with observable metrics of time and accuracy, to provide directly quantified data. We also measure user attitudes through a standardized 5-point rating scale with scores assigned to each intelligent assistant to obtain feedback from each participant on less tangible factors. Although actual quantification of these more qualified results is less trustworthy, it provides supporting evidence that we could not otherwise directly measure. For example, perhaps the tone of voice and cadence of speech may be more natural or pleasing from one digital assistant to another. These factors would only show up in a

measurement of attitude using individualized ratings.

Challenges of Imperfect Speech Recognition

Imperfect speech recognition is a major obstacle to the success of voice interface adoption. Even in a single language, there are a very large number of regional and international variations in sound for many common words. Training a digital assistant to understand many languages and many dialects and sub-dialects is an extremely complex problem. Most current products focus only on the English language.

In 2003, Bousquet-Vernhettes et al. described the incoherent remarks made by participants in an experiment of human-to-human dialogue in air traffic control after being misinterpreted which included sounds of confusion, frustration or dead silence. These recognition errors negatively affect the performance measures of speed, accuracy and overall user satisfaction with the system [2].

An important challenge related to imperfect speech recognition involves voice recognition interfaces that unintentionally discriminate against people with accents. In 2008, Zhang described how the issue of drastic speech recognition accuracy decreases for non-native speakers. This includes people who have heavy accents or who are starting to learn a foreign language. This contributes to errors in speech recognition leading to an unacceptable division between people for whom voice interfaces work [9]. Recognition of a speaker's accent before automatic speech recognition helps in improving performance [13]. A study conducted by Petkar confirms that "compared to native speech recognition, performance degrades when recognizing accented speech and non-native speech" [18].

Our findings concur with those of other researchers related to the problems of accent in speech recognition. As Huang and Deng pointed out in the 2009 work "*An Overview of Modern Speech Recognition*", "speech characteristics vary widely among speakers due to many factors, including speaker physiology, speaker style, and accents—both regional and nonnative"[7]. No new methods for overcoming this problem seem to have developed for training voice interactive systems to be more resilient to variations in speech in the past decade.

Existing voice interaction systems focus on pronunciation of native speakers of the English language. These systems are primarily trained with speech data from various native speakers of the language, which can capture and manage physiological differences, but is of no help to non-natives and those with less common accents.

The acoustic environment where voice interactions take place can also drastically influence the success of interchanges and the satisfaction of consumers. Background noise, room reverberation and overlapping speech all have profound effects on the use of spoken commands.

The success of a speech recognition module is vital to usability and if voice user interfaces do not understand the request, they are useless. This leads to a situation where the systems used to evaluate performance and improvement break down, as those consumers who require the most adaptation to allow use, will simply not use the product. Manufacturers will be unable collect the required data

and direct consumer feedback they will need most to improve their interface.

Factors that could easily perturb a consumer's satisfaction center in large part around errors in the voice recognition process. These errors disrupt an interface's ability to understand the spoken dialogue and therefore properly perform the task leading to both user frustration and failure to complete the task. Environmental interferences and competing verbal dialogue can further disrupt the assistant's voice recognition process. We used a quiet room for baseline testing, with only one participant speaking at a time.

Our findings showed a significant reduction in understanding of the request being made and therefore in accuracy of response and consumer satisfaction that was directly related to the accent of the requester's speech.

The User Study

We conducted experiments to collect information regarding the accuracy, speed and effectiveness of the queries asked to each intelligent assistant. We first describe our participants and the tasks they performed. Then we discuss the procedures used, before explaining the results.

The Participants

The study included 22 participants, recruited in the University of Washington. All of these students were university students majoring in Information Technology located in United States. Approximately 23% of the participants were female while the rest were male. The age average of this experiment group was 23.45. About 54.5% of the participants were native English speakers while the rest had a different primary language including Chinese, Spanish, Korean, Somali, and Punjabi.

The Tasks

All the participants were asked to query each of the three intelligent assistants a set of 20 questions from four different categories which included Translation, General Knowledge, "Current/Real Time Events" and "How to" questions. Five different questions were asked from each category.

Procedures

To begin the experiment, we instructed the participants to read the instructions and all of the twenty queries to be asked of each intelligent assistant. We also taught contributors how to utilize each intelligent assistant in order to avoid any issues regarding engaging the interface. After they were done with the process for each intelligent assistant they were asked to rank their user experience from 1-5 (Likert scale) and provide a brief description of why they provided the assistants with that score. We decided not to ask specific questions regarding how satisfied

they were by each intelligent assistant as we wanted to leave it to the user's own interpretation. This would allow us to compare individual satisfaction data with accuracy and elapsed time, so we could extrapolate data from their rankings and comments regarding their user satisfaction. The average experiment time per participant was around 25 minutes.

Results

Criteria factors for success used in this experiment were accuracy and speed. We measure accuracy by identifying the degree to which a question answered is correct and measure speed by the rate the question is answered in seconds. The results clearly show Cortana beat the two competitors with an impeccable accuracy of 100% followed by Alexa with an average accuracy of 74% and Siri with a mediocre accuracy of 49.8%. However, in terms of speed Siri led the race with an average speed of 2.09 seconds followed by Cortana with an average speed of 2.35 seconds and Alexa with an average speed of 2.63 seconds. Speed counts only for questions accurately answered. Questions that received 0% for accuracy are not included in the total average speed calculations. In order to objectively and comprehensively compare the different intelligent assistants, the average accuracy (percentage) was divided by the average speed (seconds) of each assistant giving a new calculation of effectiveness. The "effectiveness metric" ranked the assistants from highest to lowest as follows: Cortana (42.5%), Alexa (28.1%), Siri (23.8%).

We do not intend to imply that a difference of half a second in response time is significant by itself, despite the fact that the effectiveness metric assigns a weighting that depends on speed. Speed of response is a factor, but correct response is a more important factor. Even though Siri and Alexa have similar effectiveness metrics, Alexa has a much higher accuracy rating than Siri. In terms of usability and user experience feedback, participants scored each intelligent assistant using the Likert scale, ranking them from 1 to 5. Cortana received the highest score with an average of 4.79, followed by Alexa with a score of 3.56 and Siri with a score of 2.59. We also take explicit written feedback given by the participants at the end of the experiment into consideration when determining overall usability.

Individual Results

Siri

Siri performed the weakest in the General Knowledge and "How to" question categories. It was only able to answer an average of approximately one question per participant correctly in the "How To" category and partially answered an average of three questions correctly in the General Knowledge category. Many of these errors are due to Siri's lack of understanding of the queries asked. Since more than half of our participants' primary language was not English, we are able to extrapolate that Siri was not able to adequately recognize their accents, leading to a failure to answer questions correctly.

These results illustrate one of the most important insights regarding the accuracy and efficiency of intelligent assistants. The errors in speech recognition contribute most directly to the intelligent assistants' usability and consumer satisfaction scores. In our post query observations, one participant attested that: "Siri could not understand me half of the time; it constantly mistook the word "war" with "word" Frustrating." This was a very common theme in the evaluation of user experience.

Additionally, Siri also performed poorly due to a lack of third party application integration that might otherwise allow the application to comprehend what was being asked. Many participant questions were vaguely answered by redirecting the user to look up the information online. These questions were given an accuracy rate of 0% as the assistant was not able to appropriately answer the proposed queries. However, it should also be noted that some of the general questions and current events were answered in a quick and accurate manner. Questions from the categories of "Current/Real Time Events" and "Translation" that had a straight and dry answer were Siri's forte. Unfortunately, its lack of integration with other apps that are non-Apple hinders its effectiveness as an intelligent assistant.

Cortana

Cortana performed very well in all four categories. It gathered information from various sources and aimed to deliver in a fast manner. Cortana had the highest accuracy rate as well as the highest effectiveness metric. The cross-platform approach and integration with other applications improved the overall quality and precision of results. Cortana was able to gather results from the web as well as condense and rephrase the results to provide consumers with straightforward results, as opposed to offering alternative recommendations to finding these results.

Cortana received the highest Likert scale rating (4.75) and the feedback received was mostly positive. A participant stated: "[Cortana] exceeded my expectations. The device gave great feedback indicating it accurately heard what I said and was also able to answer all of the questions asked unlike other devices". Cortana was also able to understand all of the participant's queries and had no

problems understanding their accents as opposed to the other two intelligent assistants. Although Siri was the fastest answering queries, Cortana delivered all of its answers accurately and at a proper speed.

Alexa

Alexa performed relatively well in terms of accuracy and speed. It performed weakest in the "How to" category and took a little longer than other intelligent assistants to answer the queries. However, its third-party integrations (also known as Alexa skills) definitely helped overshadow its fellow competitors. Specific questions were answered with the help of these third-party applications especially in the "Translation" category where each question was answered correctly and even pronunciation was provided as well. However, Alexa was limited when it came to answering current event questions and was quick to apologize for not knowing the answer instead of providing other options to the user.

Alexa received a Likert scale rating of 3.56, which is moderately good. One of the participants described Alexa as "great overall but takes too long at times". In terms of query understanding, Alexa could comprehend most but not all of the queries asked. At times, it struggled to understand specific words asked by the non-native English speaker participants ending in an apologetic "I'm sorry, I don't understand". Had Alexa understood these participant's queries, it would have received an even higher accuracy rate, comparable to Cortana.

Discussion and Conclusions

This research investigated the current state of three commercially available intelligent assistants and evaluated their performance by testing voice activated queries in four different categories (Translation, General Knowledge, Current/ Real Time Events, and "How To"). We evaluated the strengths and shortcomings of these technologies through consumer testing to determine user experience with each. We undertook this work to demonstrate the enormous future capacity of voice-activated interfaces, which can be a potential aid to many other technologies.

Since this technological topic is relatively new, there is not a substantial body of related research on intelligent assistants. Our hope is that this paper can be an aid to future researchers interested in evaluating user satisfaction, which is essential for the future development of effective intelligent assistants. This work can also help conduct future research to improve the weaknesses of these interfaces. Our work did not evaluate all of the available functional tasks for each assistant, (for example having intelligent assistants make telephone calls, or make use of other related and potentially integrated applications). We instead focused on studying the efficiency of voice queries rather than cross-functional tasks that are highly personalized and contextualized to the user.

We used the data collected in this study to determine which intelligent assistant is most efficient. We measured accuracy and speed as well as usability

and feedback to evaluate the assistants. We used an effectiveness metric ranking to determine the efficiency of each intelligent assistant. We collected the participant's written evaluation of satisfaction with each assistant and they ranked each of the voice-activated assistants from 1-5.

Our experimental results showed that Microsoft's Cortana was clearly preferred over the two competitors with an impeccable accuracy of 100%. This was followed by Amazon's Alexa, with an average accuracy of 74% and Apple's Siri with only 49.8%. However, in terms of speed Siri was fastest to respond on the questions it correctly answered with an average of 2.09 seconds, followed by Cortana with an average speed of 2.35 seconds and Alexa at the average speed of 2.63 seconds. Cortana had the highest effectiveness metric, usability, and feedback scores proving to be the superior choice out of the three.

Cortana's accuracy and overall effectiveness is clearly attributable to Microsoft's unimpeachable natural language processing skills that could understand each of the participant's queries very clearly. On the other hand, Siri was unable to answer almost half of the queries asked and struggled to understand and provide a concise answer to participants' questions. Our experimental results showed that factors such as the consumer's first language and accent dramatically affect an intelligent assistants' score. Siri, in particular, could not understand several queries asked by the participants whose first language was not English. This provides a particular area for improvement as its speech recognition is currently limited to fluent English speakers.

Evaluating which category each assistant performed best (or worst) can be a strong predictor of user satisfaction which is important for the future development of effective intelligent assistants [13, 17, 18]. As seen with the results presented, the most effective intelligent assistants were the ones that were able to understand the queries when asked in a straightforward manner.

It is important to note that synonyms and ambiguities could potentially pose difficulties to these voice activated user interfaces. There are many ways people express ideas in any language. The same idea is often verbally expressed in several different ways. This ultimately forces developers to decide between accurate response and overall performance. By reducing these difficult elements of speech recognition, the interface performance can be optimized. Clearly, the most important part of answering a question, is understanding the question. Users have little to no patience for interfaces that do not understand what they are saying so matching the voice user interface with a user's mental model is crucial as it leads to higher user satisfaction and efficiency. Further work remains on the important element of speech recognition.

Future Work

Our work will be extended through the Tacoma Livable City Year Project to improve access to City services. This project is the vision of the City of Tacoma Information Technology Division. Many of the City's services are accessible by calling 311, using the Tacoma FIRST 311 mobile app, and City website. With

more and more homes having voice-activated assistants, such as the Amazon Echo (commonly called Alexa), Cortana and Google Home devices, we see the opportunity for City residents using these systems to engage them for City services. We will deploy voice-activated intelligent assistants in public places like police substations, public transit stops, libraries, shelters, etc. to give community members real time access to Tacoma city services and information. We will be monitoring these locations to determine how effective they are in satisfying consumer demand for city services and information.

We are hopeful that digital assistants will perform well in this capacity. If they prove themselves valuable, we intend to provide a model for other municipalities to expand customer services through voice interaction interfacing to information and services.

Advanced voice technology will soon be universal as natural and intelligent user interface technology integrates effortlessly into daily life. Voice user interfaces provide a natural means for users to talk to an application using conversational language. The potential future developments in the use of intelligent voice assistants are unrestrained and extensive. We look forward to continued expansion of voice interaction technologies and hope to guide future development to gradually expand the range of verbal cues to include every spoken language and dialect actively being used.

References

[1] Allan, J. and R. Jones. “Predicting searcher frustration”. In SIGIR, p. 34-41, 2010.

[2] Bousquet-Vernhettes, Caroline and Vigouroux, Nadine (2003): "Recognition error handling by the speech understanding system to improve spoken dialoguesystems". In EHSD-2003, 113-118.

[3] Chen, Rudnicky et al. “Leveraging behavioral patterns of mobile applications for personalized spoken language understanding”. In Proceedings of 2015 International Conference on Multimodal Interaction (ICMI) (2015).

[4] Chinn, John P. “Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface.” Human Computer Interaction, dl.acm.org/citation. cfm?id=57203. Originally published in International Journal of Online Engineering.

[5] Dautenhahn, Kirsten, et al. Socially Intelligent Agents: Creating Relationships with Computers and Robots. E-book, Kluwer Academic Publishers, 2001.

[6] Dong, Anlei, Yi Chang, Zhaohui Zheng, “Towards recency ranking in web search”. In WSDM, p. 11-20, 2010.

[7] Huang, Xuedong, and Li Deng. “An Overview of Modern Speech Recognition.” *Handbook of Natural Language Processing*.

[8] Jiang, Jiepu et al. “Automatic Online Evaluation of Intelligent Assistants.” ACM Digital Library, 2015, dl.acm.org/citation.cfm?id=2741669.

[9] Jiang, Jiepu, A. H. Awadallah, and R. W. White. “Understanding and predicting graded search satisfaction”. In WSDM, 2015.

[10] Kelly, Diane. “Methods for Evaluating Interactive Information Retrieval Systems with Users.” ACM Digital Library, 2009, dl.acm.org/citation.cfm?id=1618302.

[11] Kiseleva, Julia, and Kyle Williams. “Predicting User Satisfaction with Intelligent Assistants.” ACM Digital Library, 2016, dl.acm.org/citation.cfm?id=2911521.

[12] Mahmud, Khaled et al. "Voice Activated Electronic Devices Control System For Home Appliances." Scholars Journal of Engineering and Technology, digital ed., 2015, p. 66-70.
[13] Manaris, Bill. "Natural Language Processing: A Human-Computer Interaction Perspective." Advances in Computers, vol. 47, 1998, p. 1-66, www.sciencedirect. com/science/article/pii/S0065245808606658.
[14] McTear, Michael, et al. The Conversational Interface: Talking to Smart Devices. E-book, Springer International Publishing, 2016.
[15] Medicherla, Harsha, and Ali Sekmen. "Human–robot interaction via voice-controllable intelligent user interface." Robotica, vol. 25, no. 5, Sept. 2007, doi.org/ 10.1017/S0263574707003414.
[16] Minock, Michael. "Covering the Semantically Tractable Questions." ACL, 2017, www.aclweb.org/anthology/E/E17/E17-3.pdf#page=45.
[17] Mortensen, Ditte. "How to Design Voice User Interfaces." Interaction Design Foundation, Mar. 2017, www.interaction-design.org/literature/article/how-to-design-voice-user-interfaces.
[18] Petkar, Harshalata. "A Review of Challenges in Automatic Speech Recognition." *International Journal of Computer Applications*, vol. 151.
[19] Rudnicky, Alexander et al. Situated Dialog in Speech-Based Human-Computer Interaction. E-book, Springer International Publishing, 2015.
[20] Rudnicky, Alexander and Franz, A. M. "Spoken language recognition in an office management domain". In Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP), IEEE (1991), 829-832.
[21] Sproull, Lee. "When the Interface Is a Face." Human-Computer Interaction, digital ed., vol. 11, no. 2, 1996.
[22] Sun, Ming, and Alexander I. Rudnicky. "An Intelligent Assistant for High-Level Task Understanding." Proceedings of the 21st International Conference on Intelligent User Interfaces, 7 Mar. 2016, p. 169-74.
[23] Weeratunga, A. M. and SA. U. Jayawardana. "Project Nethra - An Intelligent Assistant for the Visually Disabled to Interact with Internet Services." Industrial and Information Systems (ICIIS) 2015 Conference Proceeding, 2016.