Usability of Automatic Speech Recognition Systems for Individuals with Speech Disorders:

Past, Present, Future, and A Proposed Model

Madeline Jefferson

University of Minnesota Twin-Cities

Keywords: *Inclusive design, usability, user experience, voice assistant, speech disorders, dysarthria, human factors, human-computer interaction, automatic speech recognition, voice interfaces*

Abstract

People are using voice assistants (VAs) such as Siri & Alexa more than ever before. With 46% of U.S. adults using VAs, commercially available voice-activated technologies are becoming pervasive in our homes and beyond (Pew Research, 2017). VAs provide convenience, novelty, and unique solutions for the medical industry. But, some users may be left out of the conversation. People with speech disorders or atypical speech historically have found difficulty with using automatic speech recognition (ASR) technologies, the precursor to VAs. Usability testing for these systems has consistently shown that they are not easy to use for people with speech disorders. This investigation sought to perform a literature review of the existing research on the usability of commercially available ASRs for people with speech disorders to provide historical perspectives and to take an inventory of how this issue is being addressed today.

A literature review was performed on the usability of commercially available ASRs for people with speech disorders and was divided into two stages: studies before the introduction of VAs and those that tested VAs themselves. Understanding where we have been and where we are now will also inform technical communication and usability professionals on what the future of ASRs may hold and how we can best address the needs of this audience. To do so, this paper proposes solutions for inclusive design in the voice assistant design space including a conceptual model for integrating specific techniques into commercially available VAs.

Introduction

“Alexa”, “OK, Google”, and “Hey, Siri”. These are all some of the common phrases being exchanged with today’s most popular voice assistants (VAs). Advances in voice recognition and natural language processing have made it easy to integrate speech recognition into our smartphones, smart watches, smart home devices, and smart speakers. Research conducted by Pew Research Center estimated that almost half (46%) of U.S. adults use a voice assistant (Pew, 2017). More research conducted by NPR and Edison Research found that approximately 43 million American adults own a smart speaker (The Smart Audio Report, Winter 2018. (2018)). Today, people are using this technology to perform a multitude of everyday tasks in their homes like playing the news, setting timers, home automation, and more.

But while VAs create unique solutions for many, they may not be accessible for all users. One particular community of people who are likely to have difficulty using VAs are individuals with a disorder or disability affecting their speech. Approximately 9.4 million adults in the United States have trouble using their voices (Moore et al., 2018). Often, VAs cannot accurately recognize the speech of individuals with speech disorders such as dysarthria, a motor speech disorder most commonly occurring in those who are affected by conditions such as amyotrophic lateral sclerosis (ALS), stroke, cerebral palsy, or Parkinson’s disease. Historically, many of these individuals have used automatic voice recognition (ASR) technologies because they are hands free modalities to navigate technology and are great options for people with limited or affected mobility. For decades ASRs like Dragon SpeakingNaturally have made the lives of many people with disabilities easier, including those with speech disorders like dysarthria. But with the steady adoption of VAs, are these newer speech recognition technologies living up to usability needs of these communities?

Extant research has barely explored the usability/usefulness of ASRs for people with dysarthria, let alone has it investigated VAs for these individuals. Very little consensus exists on how usable ASRs have been historically and how today's VAs are performing for people with dysarthria. Additionally, few studies have sought to understand the attitudes and perceptions of people with speech disorders surrounding these technologies.

This work seeks to understand more about usability and user experience of commercially available ASRs and VAs for people with dysarthria by gathering relevant studies. To do so, I conducted a two stage narrative literature review to evaluate the extent to which these individuals can use these systems and what their perceptions and attitudes are surrounding them. This approach helps us understand the story of how this community and ASRs have interacted, how they are interacting right now, and how they may interact with this technology in the future. I established common themes from the literature review that ultimately generated a proposed design model for voice assistants that are usable for individuals with speech disorders. The research conducted in this investigation and the proposed model serve as implications for technical communicators, usability specialists, inclusive designers, and other professionals.

Definitions and Concepts

I. Defining Automatic Speech Recognition & Voice Assistants

Automatic speech recognition refers to computerized systems that directly respond to a human voice for tasks like voice-to-text dictation. ASRs are functionally an umbrella term for any technology that converts natural human speech into text or actions. Voice assistants (VAs) are one small but very significant branch of ASR technologies (See Appendix A for more through definitions)

This paper uses the term voice assistants, but it is important to note that these systems have been called many other names including: intelligent personal assistants, automated personal assistants, intelligent assistants, voice-controlled virtual assistants, voice-controlled digital assistant, smart assistants, conversational assistants and voice controllers. While there are subtle differences between each of these terms, the terminology used can define the functionality. Hoy et al. wrote that voice assistants are, "software agents that can interpret human speech and respond via synthesized voices" (2018, pp. 81). For the purpose of this investigation, the term voice assistant is used to refer to voice recognition software that meets the following criteria: 1) uses natural language input from a speaker, 2) responds with a programmed digital voice, 3) carries out predetermined actions, 4) is available in a commercially sold device.

Voice assistants are available in a wide variety of devices. Pew Research Center found that most U.S. adults use VAs on their smartphone (42%), followed by use on a computer or tablet (14%), and some (8%) use a stand-alone device such as a smart speaker (Pew Research, 2017). This literature review refers to voice assistants in all of these contexts.

While there are many proprietary voice assistants, this investigation focuses on the most popular commercially available voice assistants Apple's Siri, Google's Google Assistant, Amazon's Alexa, and Microsoft's Cortana. This investigation is focused on commercially available ASRs and VAs because they are the systems most likely to be used in daily life by the average consumer with dysarthria.

II. *Defining Speech Disorders & Dysarthria*

The term speech disorders can refer to a large range of abilities and differences. The American Speech-Language-Hearing Association defines a speech disorder as "an impairment of the articulation of speech sounds, fluency and/or voice" (Definitions of Communication Disorders and Variations, 1993). Speech disorders can affect the production, rhythm, rate, quality, pitch, loudness, and duration of speech. This includes individuals who may be diagnosed with a primary speech disorder such as stuttering or those who experience a speech disorder as a result of another diagnosis such as amyotrophic lateral sclerosis (ALS). In the context of this paper, speech disorders will be used to refer to a condition that produces speech which is atypical in volume, rate, or quality that affects the ability of the speaker to be understood.

Dysarthria, one of the most common speech disorders, is caused by some form of brain damage or abnormality and produces weakness or difficulty controlling the muscles involved in speech. The disorder is characterized by slow, quiet, and disjointed/unpredictable speech that can make an individual difficult to understand (Dysarthria, 2019). These characteristics affect the overall the percentage of speech that a listener with normal hearing can understand, also known as speech intelligibility. This disorder typically arises as a secondary condition and is common for individuals with cerebral palsy, ALS, stroke survivors, multiple sclerosis, Alzheimer's, Parkinson's, and traumatic brain injury. The following investigation focuses on individuals with dysarthria because it is one of the most prevalent speech disorders.

Background

I. The History of Automatic Speech Recognition & Voice Assistants

Before exploring this topic, it is important to understand what led up to commercially available ASRs and VAs. ASR technology has existed since the early 1950s but the technology did not experience major milestones until the 1960s. In 1962, a group of IBM engineers demonstrated the "Shoebox" at the Seattle World's Fair. This tool could perform calculations by recognizing 16 spoken words and digits 0-9 (Juang, 2004). A decade later, scientists at Carnegie Mellon University created Harpy, an ASR that was able to recognize 1,011 words, approximately the vocabulary of a three-year-old with reasonable accuracy (Juang, 2004). Advances in this area were often funded by the government or were developed for proprietary use. But, as computing power grew cheaper to manufacture and more powerful, several commercially available ASRs became available to the public. In 1997, the release of Dragon NaturallySpeaking set a precedent for ASR software with its ability to transcribe normal speech with 95% accuracy (Fortune, 1998). This system also included several features that accommodated atypical speech characteristics and quickly became widely used by people with disabilities who could not type on a computer with their hands (Ferrier, 1995).

Then, with the advent of mobile phones, systems that already had a microphone and speaker built in, speech recognition systems were able to take the next leap and voice assistants were developed. In 2011, Apple released their voice assistant Siri bundled into the software on every iPhone. For the first time ever, VAs were used by the masses in a convenient and portable way (Hoy, 2018). Early adopters of this tech could enjoy using their voice to dial a friend or to set a timer. Shortly after, Microsoft released their own VA, Cortana (2013), followed by Amazon's Alexa (2014), and Google's Assistant (2016) (Hoy, 2018). Amazon's Echo series gave way to a new era of stand-alone, in home devices for VAs. Amazon and Google were the first of the major tech companies to create smart speakers, followed shortly by Apple.

II. How ASRs & VAs Work

Voice assistants rely on automatic speech recognition technology. In essence, ASR is the process of automatically converting natural human language to a digital sequence of words. There are three categories of ASR systems: (a) speaker dependent, (b) speaker independent, (c) speaker adaptable (Young & Mihailidis, 2010). Speaker dependent ASRs require training prior to use. These systems can be trained using audio files or live recordings of speech and are typically customized to respond to one speaker. Speaker independent ASRs do not require training prior to use and can respond to multiple different speakers. Most users will be able to interact with a speaker independent ASR easily as long as their speech falls within a predetermined range. Speaker adaptable ASRs require no prior training and adapt to the user over time. Typically as a user interacts with this kind of ASR, the system uses machine leaning to refine accuracy.

Voice Assistants, a branch of ASR technology, use spoken language as both the input and the output of the system. To begin a speech input, most VAs have a "wake word" or phrase to signal to the device that it is time to start listening for a command. In the case of Apple's Siri, this phrase is "Hey Siri", for Amazon's Alexa, "Alexa", for Google Assistant, "Ok Google", and for Microsoft's Cortana, "Hey Cortana". Once a speaker is done giving a command or asking a question, the speech recording undergoes processing and signal optimization. Speech recognition algorithms convert speech into text which is used to generate a response from the system. The VA then responds to the query with a computerized voice (Hura, 2017). Depending on the command, the VA may also carry out an action, play requested media, or complete a task with a variety of services and devices. Most commercially available VAs also store these recordings in some form to train built in machine learning algorithms that analyze details of the users speech over time. This generally makes the VA respond more accurately to the user with time and use. Most VAs use a combination of speaker independent and speaker adaptable ASR technology.

III. Relevant Work

VAs are useful for people with disabilities for many reasons, primarily their hands-free nature. Pradhan et al's paper, "Accessibility Came by Accident", states that, "from an accessible technology perspective, [VAs] offer the potential to apply speech input and output beyond the traditional confines of text dictation and screen reader software" (Pradhan, 2018, pp 1). While we are just beginning to understand the uses of this technology, we know that many individuals with disabilities are already benefitting from the unique options VAs have to offer. For example, a person with limited mobility can use a VA to control the the lighting in their house, while a user with vision impairment can ask for the weather without having to navigate a screen reader. Coyne et al. conducted an analysis of reviews of Amazon's Echo and Dot devices and found that users with disabilities, particularly those with limited mobility reported an improved quality of life with the use of this VA (2017).

The increased independence VAs can provide for their users is even beginning to be recognized as a unique opportunity for several populations. This year, Cedars-Sinai Hospital Los Angeles is piloting a program placing smart speakers in patient rooms to give them more independence and to free up nurse time for critical cares (Shu, 2019). Today, VA tools are being designed for users with unique needs including, cognitive support during daily tasks for users with dementia and Alzheimer's (Carroll et al., 2017), automatic emotion detection and support for users with neurodevelopmental disorders (Catania et al., 2019), and support for users with mild TBI and PTSD (Wallace & Morris, 2018).

Unfortunately, all of these benefits rely on the ability of the system to recognize and properly respond to a user's speech, an issue that heavily affects the usability of ASRs and VAs for individuals with dysarthria. Young and Mihalidis write that, "Current commercial speech-to-text…systems are designed specifically for a mainstream, non-speech-disordered adult population, thus purposely excluding individuals with speaking disorders" (2010, pp. 26).

METHODS

To understand how individuals with dysarthria are interacting with speech recognition technology today, this literature review seeks to show the history of the usability of ASRs and to take an inventory on the current research of the usability of VAs for this population. Understanding the historical perspectives of commercially available ASR usability informs us about the usability of the most popular VAs on the market today: Siri, Alexa, Google Assistant, & Cortana. To achieve this, I gathered a collection of peer-reviewed literature and analyzed it to explore three questions:

1) How usable were previous commercially available ASRs for individuals with dysarthria?
2) How usable are popular VAs today for individuals with dysarthria?
3) What are the attitudes and perceptions of these individuals about the technology?

Extant research does not contain a literature review on this topic nor does it adequately discuss the attitudes and perceptions of these users. Thus, this work aims to generate further insights on the current state of VAs and to understand the historical perspectives of ASRs for these individuals to drive inclusive design for this technology.

Narrative Literature Review

I performed a literature review in two parts with Part I pertaining to research on the commercially available ASRs from 1990 to 2011, and Part II pertaining to the usability of commercially available VAs from 2011 to the present for people with dysarthria. The year 2011 serves as a cut off between the two searches because it is the year that the first commercially available VA, Apple's Siri, was introduced. To perform the literature review, I searched several databases for articles: ResearchGate, Web of Science, PubMed, IEEE, and Google Scholar.

To locate relevant articles, I defined two sets of specific keywords for each search (See Table 1 & 2). Keywords were broken into three categories: system, diagnosis, & method. For every search, I used one keyword from each category and results were analyzed for relevancy. This process was completed for both Part II & II on the five databases mentioned above.

TABLE 1
Part I: ASRs 1990-2011

Category	Keywords
System	Automatic speech recognition, speech recognition, speech recognizer, voice control, dictation, Dragon Dictate, Dragon SpeakingNaturally, VoiceType, VoicePad
Diagnosis	speech disorder, speech impairment, speech disability, disability, dysarthria
Method	usability testing, reviews, user interaction, user testing, survey, interview

TABLE 2
Part II: VAs 2011-Present

Category	Keywords
System	voice assistant, intelligent personal assistant, automated personal assistant, intelligent assistant, voice-controlled virtual assistant, voice-controlled digital assistant, smart assistant, Siri, Alexa, Cortana, Google Assistant
Diagnosis	speech disorder, speech impairment, speech disability, disability, dysarthria
Method	usability testing, reviews, user interaction, user testing, survey, interview

From the search results, I chose studies for analysis if they met the following criteria:

1) Were published in a peer-reviewed journal or were submitted to a professional conference from 1990-present
2) Studied the usability or user perspectives of an off-the-shelf ASR or VA
3) Worked with participants with either a self-identified or a formally diagnosed speech disorder or dysarthria
4) Evaluated performance of an ASR or VA by using rates of correct word/speech recognition, word error rate, or correct response rate.

Results

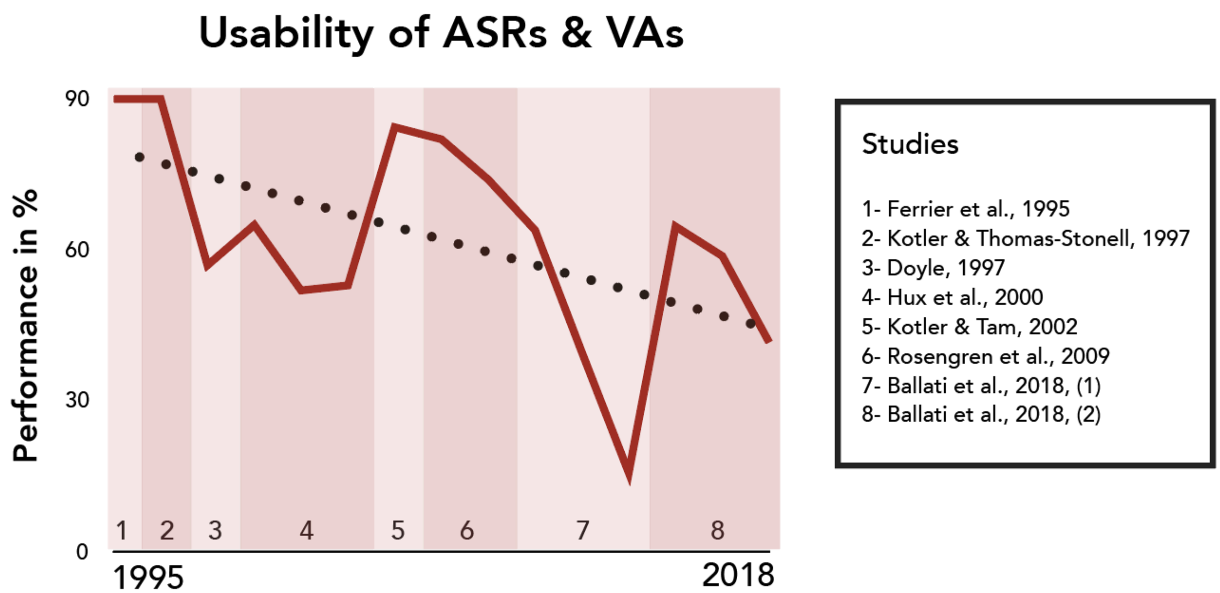
Using the exclusion criteria above, this investigation uncovered a total of 12 articles relevant to the research question. Of this, six articles were used for Part II of the literature review and four for Part II. Two of the articles were excluded from the review because they were literature reviews but were still used for context and conclusions. Results indicate that ASRs before the era of VAs were indeed easier to use for individuals with dysarthria. In Figure 1, we can clearly see that over time, these systems have tested poorly for user with dysarthria.

Of the six studies identified for Part I, four tested either Dragon Dictate or Dragon SpeakingNaturally, three tested VoiceType, one tested VoicePad Premium, and one tested Microsoft Dictation. Only one study was identified that explored the user experience of these systems for people with dysarthria. Of the four studies used for Part II, three tested Google Assistant, two tested Alexa, two tested Siri, and only one tested Cortana. Only one study explored the user experience of VAs by analyzing Amazon reviews (See Appendix B for a summary of all articles reviewed in this investigation).

Performance of ASRs were primarily evaluated using speech recognition rate: a measure of the accuracy of a system to convert speech into text. For ASRs, a speech recognition rate between 90-95% is considered satisfactory for people with non-disordered speech (Rosen & Yampolsky, 2000). Some researchers have considered an

80% speech recognition rate satisfactory for dysarthric speech (Ferrier et al., 1995) while others consider 65% “quite acceptable” for this population (Hux et al., 2000). Many studies identified in this review also examined word error rate: a measure of how frequently a speech recognition system has made an error in transcription. While there is no major consensus on what is an acceptable word error rate, logically, lower error rates are preferable. Voice assistant performance was primarily evaluated by the ability of the system to correctly respond to certain commands or comments, also known as correct response rate. Again, I was unable to locate an industry standard for acceptable correct response rate.

FIGURE 1



Part I: ASRs 1990-2011

Many of the studies on ASRs found a surprisingly high speech recognition rate for individuals with a variety of degrees of dysarthria. Usability work conducted by Ferrier et al. in 1995 tested the efficacy of DragonDictate by measuring the the number of sessions it would take for participants with dysarthria to reach at least 80% speech recognition. The researchers assumed that only participants with mild to moderate speech would be able successful with this system but found that all participants were able to achieve a recognition rate of 90% or higher after only using the system in 8 sessions. In 1994, Rosengren et al. compared the performance of DragonDictate and Infovox for individuals with dysarthria and found average speech recognition rate to be 82% and 74% respectively (2009). Some research also suggests that users can undergo specific speech therapy to use these systems with more ease. In a case study by Kotler and Thomas-Stonell, one participant with mild dysarthria underwent 12 sessions to improve specific speech sounds and his speech recognition improved from 72% to 90% when using IBM’s VoiceType program (1997).

Still, other studies found these systems performed poorly for this audience regardless of their level of dysarthria or speech intelligibility. Research conducted by Kotler and Tam (2002) tested DragonDictate & VoiceType on speakers with dysarthria and found the average speech recognition to be 62.4-84.4%. One participant with mild dysarthria had an average speech recognition rate as low as 52.6%. More work conducted by Hux et al. (2000) tested the speech of one participant with dysarthria on three commercially available ASRs: Microsoft Dictation, DragonNaturally Speaking 3.0, VoicePad Premium. After 5 sessions of training, they found that the participant, despite having a speech intelligibility of 98.18%, was only able to achieve a nearly 65% speech recognition rate for DragonDictate, 52% for Microsoft Dictate, and 53% for VoicePad. Another study testing VoiceType conducted by Doyle et al, found that recognition rates did vary by severity but that even speakers with mild dysarthria were only able to achieve a 57% speech recognition rate (1997).

Only one study was identified in Part I of the review that explored user perceptions and attitudes of ASRs pre 2011. The Kotler and Tam investigation interviewed participants on their perspectives of Dragon Dictate and VoiceType and found that users were satisfied with the hands-free nature of the systems because they were faster than previous hands-free modalities for typing and that these ASRs increased their productivity (2002). Despite this, many participants reported the systems were still too slow and required a significant amount more time to correct mistakes when compared to control participants. Some users also reported frustration, fatigue, discomfort, and feeling self conscious when using the systems.

Part II: VAs 2011-present

Two of the studies of importance to this investigation, conducted at the Polytechnic University of Turin, researched the usability of voice assistants for people with ALS. In one study, Ballati et al. (1, 2018) found that users with moderate dysarthria were able to achieve about a 60% correct response rate for both Google Assistant & Siri and a very poor 25% for Cortana. Another study conducted by Ballati et al. (2, 2018) testing Alexa, Siri, & Google Assistant found Google Assistant had the highest correct response rate of 58.82%, Siri had 47%, and Alexa had a correct response rate of 41.17%.

More work in this area conducted by Moore et al. compared the word error rate of Google Assistant & CMU's Sphinx when being used by participants with dysarthria and found that Google Assistant had a significantly lower word error rate (2018). The Sphinx system had an 84% higher word error rate than Google Assistant in this study.

Finally, the only study in Part II that explored the attitudes and perceptions of users on VAs was Pradhan et al's "Accessibility Came By Accident: Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities" (2018). Their investigation consisted of doing a content analysis of reviews of Amazon's Echo, Echo Dot, & Tap smart speakers to identify themes from users with disabilities. While this study did not focus on users with dysarthria specifically, 30.6% of the reviews they analyzed were from people with motor or mobility impairments & 13.6% were from users with a speech impairment. A common theme reported in the reviews was ease of use even for users who disclosed having a speech disorder. Over 74% of user reviews that mentioned having a speech or voice impairment were positive. Some of the positive reviews from users with speech impairments stated, “Ordinarily voice programs can’t understand what I am saying due to my speech impairment, but Alexa responds to my commands without fail” and “Most humans can not understand me, but Alexa can” (Pradhan et al., 2018, pp. 6). Positive reviews also called out the independence VAs gave them with reviews like this one,

“I can’t begin to tell you what a difference the echo has made to my disabled veteran husband. After his stroke, his mobility and speech were effected. Giving him a whisper of a voice. He can now ask Alexa to play any song of his choosing without having to getup.” (Pradhan et al., 2018, pp. 4)

Nevertheless, 10 reviews from users with speech impairments mentioned difficulties with using the system due to their atypical speech.

Discussion

While the research on the usability of commercially available ASRs & VAs for this group is limited, it still overwhelmingly suggests that these systems continue to fail to perform to industry standards for speech recognition for this community. Previously commercially available ASRs (Part I) were generally found to be usable for this group with recognition rates ranging 52% to 90%. Studies in Part I did suggest that severity of dysarthria did affect the usability of these systems but that DragonDictate, even users with severe dysarthria were able to achieve at least 80% recognition. The user attitudes for studies in Part I illustrated that individuals with dysarthria were satisfied with this hands-free modality for dictation but did experience issues with it including: fatigue, feeling self-conscious while using, frustration, slowness, and “unacceptable recognition accuracy” (Kotler and Tam, 2002, pp. 141).

Today’s commercially available voice assistants pale in comparison to the usability of ASRs. Studies in Part II found correct response rates ranging between 15.81% and 63.96% indicating that usability for this technology has decreased for people with dysarthria and other speech disorders. Again, researchers found that the severity of dysarthria affected performance of the system. Results for user attitudes and perceptions for Part II were similar to findings in Part I. Overall, users has positive experiences with voice assistants but noted several issues with using them related to their speech impairment.

It is very interesting to note that the recognition rates of ASRs before the advent of VAs (pre-2011) were significantly higher than the current off-the-shelf voice assistants. From the literature review we see the story of a technology that arose as a usable tool for people with dysarthria but now functions as a one-size-fits-most tool that systematically excludes these communities. Additionally, user attitudes were more or less the same implying that these individuals want to use this technology but still experience the same problems when using it.

This may be due to the fact that most VAs are developed using a small range of speech that does not include users with atypical speech. An article published to Scientific American in 2016 notes that “Speech recognizers are targeted at the vast majority of people at that center point on a bell curve. Everyone else is on the edges” (Mullin, 2016). This effectively restricts users with dysarthria and other speech disorders from being able to reap the many benefits of VAs. Moore et al. wrote that, “This creates a social barrier which prevents some individuals with voice disorders from fully participating in their community” (Moore, 2018, pp. 469). In many ways, we have actually regressed in the usability of these systems for people with dysarthria.

Limitations & Future Research

Due to the limited number of studies on the usability and user experience of commercially available ASRs & VAs for individuals with dysarthria, this literature review had many limitations. Primarily, there are so few studies in this area that it is difficult to draw any conclusions. In conducting the literature review, plenty of experimental research testing the usability of custom designed ASRs and VAs was uncovered but very few researchers have investigated how the tech companies developing the most popular VAs are designing for people with dysarthria and speech disorders in general. Additionally, I experienced some difficulty accessing all articles on this topic because many were conference proceedings that were behind a pay wall. These articles had to be excluded from the review but may have contained findings important to this work.

Some of the limitations of this work were not with the review process but with the research itself. The studies utilized in this review did not show consistent practice for how many sessions were used to determine the efficacy of each system. Some studies only conducted one session for testing while others used up to eight. As most ASRs and VAs include some element of machine learning, speech recognition rates should theoretically improve over time. Without a consistent number of testing sessions, results are difficult to compare across studies. As each of the four commercially available VAs investigated (Siri, Alexa, Google, Cortana) utilize deep learning to tailor the system to the speaker, there should be research to uncover if these systems could improve for users with dysarthria over time.

Of the 10 studies included in the final review, none investigated longer term use of ASRs or VAs. Longitudinal studies are also critical in this area as the machine learning integrated into commercially ASRs and VAs is subject to what Young and Mihailidis (2010) refer to as “voice drifting”. Voice drifting occurs when, “the ASR system starts to adapt to altered or fatigued speech, thereby increasing the possibility of mis-recognition when non-fatigued speech is used” (Young and Mihailidis, 2018, pp. 105). In performing the literature review, one article was found by Bentley et al. that explored the long term use of smart speakers (2018). However, this research was not included in the review as the user group was not specifically users with speech disorders. Without data on long-term use of these systems, we cannot conclude the prevalence of this issue and if it is affecting the usability over time.

Furthermore, the research that does exists focuses primarily on the usability aspect of these systems and does not shed much light on the user attitudes and perceptions that surrounds them for this community. Only two studies were found to explore user experience, attitudes, or perceptions: one for the ASRs Dragon Dictate & VoiceType and another for Amazon's Alexa. Kotler and Tam's (2002) interviews of users with dysarthria provide limited insight into the experience of these individuals in their interactions with the ASRs. Meanwhile, Pradhan et al's (2018) content analysis of Amazon reviews, while an inventive approach, is wrought with limitations. The contrived context of Amazon reviews leaves out the perspectives of individuals who did not review the device online and leaves us with the extremely subjective perspectives of those who are the type of people to leave reviews. Review data is also limiting because our personas and actions can change when using the anonymity of the internet.

Very little formal research on the usability of commercially available VAs for users with speech disorders that aren't dysarthria was found during this review. Performing usability testing for people with dysarthria, while important to study, provides only a subset of information on the usability of these systems for people with speech disorders. There is some work that explores the usability of ASRs and VAs for users with speech affected by a hearing impairment (Fok et al, 2018) & for at least one study on the feasibility of using these systems for individuals who use AAC (Bryen & Chung, 2018). Further research needs to be done in this area if we are to design systems that work for all users.

Finally, because there is no consensus in the industry on what terminology to use when referring to voice assistants, it is difficult to ensure that all articles relevant to this investigation were uncovered. The defined keywords for the literature review were through but not exhaustive of all of the terms and jargon surrounding ASRs and VAs. This firmly establishes the need for universal language in regards to these technologies.

Where Do We Go From Here?

Ultimately this leads us to grapple with the idea that today's most influential tech companies, Apple, Microsoft, Google, and Amazon might not be meeting the needs of these users. Perhaps Steven Aquino, a technology blogger, said it best with the line, "I hope Apple and Amazon and other companies are investing in training Siri and her ilk to learn speech (disorders)…If voice is the future, as many in the commentariat believe it to be, then accessibility must be looked at differently." (dotCMS, 2019). If we are to ensure true inclusive design in future voice-interfaces, accessibility should not be viewed as an additional feature and must be integrated into the systems themselves.

I. How Are Big Tech Companies Addressing This Problem?

One way we can evaluate how these companies are addressing this issue is by exploring the conversations they are having with the public and with their business partners around making voice assistants accessible. To understand more on this, I searched for the accessibility mission statements and developer guidelines on the websites of each of the four major tech companies developing voice assistants today: Apple, Amazon, Google, and Microsoft. I also was also searched their respective sites for information on accessibility of each of their voice assistant specifically. Using the native search bar on each company's site, I found pages related to accessibility and each company's voice assistant and analyzed them for mention of speech disorder, impairment, or impediment.

Apple

Taking a cursory look at this documentation doesn't yield much confidence that Apple is taking this population seriously. Apple, the company with the longest running voice assistant, Siri, states on their accessibility webpage that, "We want everyone to enjoy the everyday moments that technology helps make possible, so we work to make every Apple product accessible from the very start" (Apple Accessibility). Their accessibility mission statement page outlines what specific strides they are making for users with disability by sorting achievements into categories: vision, hearing, physical and motor skills, and learning and literacy. The category they failed to address? Speech. Additionally, Apple's statements to their developers do not mention how they recommend their business partners address the needs of users with atypical or unique speech (Accessibility for Developers).

Microsoft

Microsoft's accessibility webpage also breaks the populations they are making specific accommodations for into categories: vision, hearing, neurodiversity, learning, mobility, and mental health (Microsoft accessibility). Once again, we see no mention of the steps being taken to make their technology more usable for people with speech disorders. One page from Microsoft's developer site titled *Accessibility Overview* mentions speech impairments but does not call out specific design recommendations for this community (Developing apps for accessibility).

Amazon

The company with the most explicitly outlined guidance for speech accessibility was Amazon. While their main accessibility page does not mention users with speech disorders or impairments, at least two of their developer guideline pages do. One page, Make Your Skill Accessible to All, states, "It can be difficult for customers, even those without speech impediments, to tell Alexa long strings of numbers or detailed information" and "Customers with learning or speech disabilities might find overly verbose skills particularly difficult to use" (Make Your Skill Accessible to All). Amazon even provides a page on accessibility for Alexa specifically that calls out strategies for users with speech impairments like integrating user feedback and the ability to customize wake words (Accessibility Features for Alexa).

Google

Google's accessibility page links to templates for accessible design of some of their products such as Google Drive and Gmail, but does not make any attempt to identify who they are designing for (Google Accessibility). Moreover, Google's page for accessibility guidelines for developers simply states, "Google encourages developers and publishers to design and build products and applications with accessibility in mind" and, "Making applications accessible not only ensures equal access to the roughly 1 billion people in the world with disabilities, but also benefits people without disabilities by allowing them to customize their experiences" (Resources for Developers and Publishers). Perhaps most disappointing is that the company appears to have no page dedicated to discussing the accessibly of their VA, Google Assistant.

II. How are Researchers and Independent Developers Addressing This Problem?

Researchers and developers are addressing making ASRs & VAs more usable for people with dysarthria and other speech disorders in a variety of ways: some positive, some regrettable. Some professionals like Kotler and Thomas-Stonell (1997) recommend training the user themselves for better interactions with ASR systems by using specific speech therapy techniques. This approach is not only outdated but crucially fails to achieve any form of user-centered design. Don Norman, author of the world famous design book *The Design of Everyday Things,* writes that human-centered design is, "an approach that puts human needs, capabilities, and behavior first, then designs to accommodate" (Norman, 2013). The fatal flaw in Kotler and Thomas-Stonell's approach is that they attempted to force the user to fit the needs of the system which is unethical and at the end of the day unintuitive design.

Some researchers and developers that have gotten closer to solving this problem by developing VAs and ASRs specifically for users with speech disorders. Derboven et al. (2014) developed an ASR system named ALADIN that, "adapts to the user's language, and does not impose a predefined vocabulary" (2014). This speaker dependent system was found to be very effective in responding to the speech of users with speech disorders. Researchers at the University of Edinburgh developed a creative solution for users with degenerative disorders that affect speech such as Parkinson's (Yamagishi et al., 2012). Their creative approach is to "bank" the voices of users with speech disorders and use that stored data to piece together unintelligible speech as a disease or condition progresses. In effect, machine learning systems can be taught to adapt to the changes in a users voice as their speech becomes more difficult to understand over time.

Even several independent companies are currently tackling this problem space. The company VoiceItt, a 2018 winner of Microsoft's global AI for Good award, states on their site that they are, "developing the world's first speech recognition technology designed to understand non-standard speech" ((http://www.voiceitt.com). Their technology transforms the unintelligible speech of users with a speech disorder and transforms it into a computerized voice that can be easily understand. Another tech company TECLA uses their portable hands free assistive device to connect to Amazon's Alexa. Using the Tecla app, users with difficult to understand speech can create on screen buttons on their smart device that function as commands for Alexa (Tecla App: Amazon Alexa).

While these solutions are taking important steps toward the inclusive design of VAs and ASRs, their biggest downfall is that they require extra work on the part of the user like an additional application download or pairing a proprietary device. Why is the tech industry relying on the innovations of smaller developers to drive inclusive design? This approach puts the burden of finding and implementing usable technology on the user group itself. Why are the companies with the most influence making these systems unusable for the people who could benefit from using them the most? This essay seeks to propose the following conceptual model intended to make today's most popular commercially available VAs usable for individuals with speech disorders.

Proposed Model for the Inclusive Design of Voice Assistants

Many strategies can be used to make VAs and voice interfaces in general more useable for people with atypical speech. A VA design that would properly address the needs of these users would need to be accessible with little to no set up, would account for degenerative voices or voices that drastically change with time, and would easily understand atypical speech. Developing a voice assistant that could properly function for these communities could be achieved in a variety of ways:

1) Train with Diverse Speech

We could train the VA with a more diverse range of voices. Moore et al write that, "The datasets used to train ASR systems need to be more inclusive of different voices than the current datasets" (Moore, 2018, pp. 469). This ensures that the system is more likely to respond to the unique speech characteristics of users with speech disorders like dysarthria.

2) Use Multiple Modalities

Features that use multiple modalities are more likely to understand what a user is trying to say or achieve. ASHA recommends watching speakers with dysarthria while they talk to increase comprehension (Dysarthria ASHA). A VA could include a camera function that uses facial recognition software to read lips to increase the likelihood of the system correctly responding.

3) Optimize the Signal

We could use the techniques set forth by Derborven et al. (2014) to make the VA better adapted to respond to atypical speech behaviors such as repetitions of sounds, pauses & abnormal speaking rate. Optimizing the speech signal of users with the specific computing techniques could be very effective. Benzeghiba et al. (2007) performed a literature review on specific techniques on how to enhance the recording of speech that ASRs use and provided insights into how this could be achieved. Improving the speech input that the VA uses would reduce the burden of usability for the user significantly by allowing them to continue to use their own natural voice to interact with the system.

4) Make Systems Speaker-dependent

Several researchers advocate for the use of speaker-dependent ASRs for people with speech disorders including Mulfari et al. (2018) and Neerincx et al. (2009). While this requires more set up time, it would allow the VA to understand how each user would pronounce specific words and sounds. It is likely that only users with severely affected speech would require this but it could be included as an optional part of the set up process for the VA or a feature that could be toggled on at a later date.

5) Use Machine Learning

Most commercially available VAs are already using machine learning to adapt the system to a user over time but more specific techniques for users with speech disorders could be used. Accounting for voices that change over time by “banking” speech (Yamagihi et al., 2012) and accounting for the effect of voice drifting could make machine learning more effective for these users.

6) Create Speech Disorder Friendly Content

We could ensure that the content written for the VAs are easy to use for people with speech disorders. Amazon’s developer guidelines for Alexa skills indicate that content created for VAs should not require the user to speak too many words (Make Your Skill Accessible to All). Developers and technical communicators could also try write to avoid using specific speech sounds that have been found to be difficult to pronounce for people with speech impairments.

Most importantly a VA with the above attributes would need to be embedded in the commercially available options that already exist today. Integrating this technology into the existing hardware and software that people already use is the most inclusive way to address this issue. It allows all users to access technology that works for their unique needs without a pay wall or additional financial burden to them. Doing so shifts the responsibility of making VAs usable from the consumer to the manufacturer and designer.

We know it is feasible to make VAs usable for people with speech disorders an industry standard by looking at the case study of Apple's VoiceOver screen reading technology. Screen readers are a crucial tool for people with blindness and vision impairment to navigate computers and mobile phones. With the rise of personal computing in the 90s and early 2000s, several companies released downloadable software that could read code and translate it into actions for blind users. In 2005, Apple released VoiceOver bundled into Mac OS X 10.4 Tiger making the company the first operating system vendor to build a fully functional screen reader that didn't require any additional installation procedures (Morales, 2015). Shortly after, many tech companies began following suit. Not only did this action demonstrate that Apple deeply cared about inclusive design, it increased the number of users that could benefit from their product. To this day, the majority of blind and vision impaired smartphone users use Apple's products because of their continued efforts to create usable technology for this group. This generates significant revenue. If we apply this same mentality to making VAs accessible, it's easy to see how achievable this is.

Final Thoughts

The goal of this investigation was to take a critical look at the historical and current research of the usability ASRs & VAs for users with speech disorders to construct an image of future research and design work that needs to be done in this area. Fundamentally, this literature review and proposed design model uncovered many potential future projects and demonstrate just how much work still needs to be done on this topic. It is discouraging to see that historically ASRs were easier to use for people with speech disorders than today's VAs are. It appears that tech companies may have spent so much time focusing on creating VAs that work out of the box that they have neglected that some users need customization options. They have failed to speak to how they are addressing the needs of users with speech disorders and do not make their expectations for voice accessibility for their partners in development clear.

Fortunately, we can see from the proposed model that there are many ways to address this problem that are feasible to implement. The reality that ASRs and VAs frequently perform poorly for users with speech disorders is fundamentally a human factors issue. Leaders in the usability and user experience industry would agree that human factors problems arise when elements of a design prevent the use of a system or product by a specific audience. That is exactly what we see at work here. If we view this problem as a human factors issue, we can easily begin to think about how best to design for this user group.

It's important to note that our interactions with voice assistants are not only convenient, they are meaningful and similar to how we interact with other humans. Research on the human-computer interaction of VAs on mobile phones found that although participants knew the voice they were hearing was being produced by the VA, they still thought of it as, "representing the mobile phone" (Guzman et al., 2019). Participants in this study were found to heavily personify their built in VA rather than view it as a computerized tool. This makes VAs unique and highly emotional technologies that we need to learn more about.

Because this technology is starting to be used in a wide variety of contexts, we need to understand how to make it accessible to all users including the content that technical communicators will create for it. While the onus of creating VAs that are usable should fall on the large technology companies driving the innovations of these systems, professionals in the usability, technical communication, and design industry should be aware of this problem and the specific ways we can address it. Only with this awareness will be able to truly create usable VAs for people with speech disorders.

References

Apple. Accessibility. Retrieved from https://www.apple.com/accessibility/

Accessibility Features for Alexa. Retrieved from https://www.amazon.com/gp/help/customer/display.html?nodeId=202158280

Apple Inc. Accessibility for Developers. Retrieved from https://developer.apple.com/accessibility/

Ballati, F., Corno, F., & De Russis, L. 1 (2018). Assessing Virtual Assistant Capabilities with Italian Dysarthric Speech. *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '18*, 93–101.

Ballati, F., Corno, F., & Russis, L. D. 2 (2018). *"Hey Siri, do you understand me?": Virtual Assistants and Dysarthria*. 11.

Bentley, F., Luvogt, C., Silverman, M., Wirasinghe, R., White, B., & Lottrjdge, D. (2018). Understanding the Long-Term Use of Smart Speaker Assistants. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *2*(3), 1–24.

Benzeghiba, M., De Mori, R., Deroo, O., Dupont, S., Erbes, T., Jouvet, D., … Wellekens, C. (2007). Automatic speech recognition and speech variability: A review. *Speech Communication*, *49*(10–11), 763–786.

Bryen, D. N., & Chung, Y. (2018). *What Adults Who Use AAC Say About Their Use of Mainstream Mobile Technologies*. *12*, 34.

Carroll, C., Chiodo, C., Lin, A. X., Nidever, M., & Prathipati, J. (2017). Robin: Enabling Independence For Individuals With Cognitive Disabilities Using Voice Assistive

Technology. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17*, 46–53.

Catania, F., Nardo, N. D., Garzotto, F., & Occhiuto, D. (2019). *Emoty: An Emotionally Sensitive Conversational Agent for People with Neurodevelopmental Disorders*. 10.

Coyne, M., Thomas, C., Collimore, A., Franzese, C., & Hwang, C. (2017). Early User Centered Insights on Voice Integrated Technologies Through Retrospective Analysis. *Iproceedings*, *3*(1), e49.

Definitions of Communication Disorders and Variations. ASHA. (1993, January 01). Retrieved from https://www.asha.org/policy/rp1993-00208/

Derboven, J., Huyghe, J., & De Grooff, D. (2014). Designing voice interaction for people with physical and speech impairments. *Proceedings of the 8th Nordic Conference on Human-Computer Interaction Fun, Fast, Foundational - NordiCHI '14*, 217–226.

Developing apps for accessibility. Retrieved from https://developer.microsoft.com/en-us/windows/accessible-apps

dotCMS. (2019) Smart Speakers & Web Accessibility: How Voice Technology is Helping Disabled People. Retrieved March 20, 2019, from dotCMS Content Management System website: https://dotcms.com/blog/post/smart-speakers-web-accessibility-how-voice-technology-is-helping-disabled-people

Doyle, P. C. (1997). Dysarthric speech : A comparison of computerized speech recognition and listener intelligibility. *Journal of Rehabilitation Research and Development, 34*(3), 309-316.

Dysarthria. ASHA. Retrieved April 15, 2019, from American Speech-Language-Hearing Association website: https://www.asha.org/public/speech/disorders/dysarthria/

Ferrier, L., Shane, H., Ballard, H., Carpenter, T., & Benoit, A. (1995). Dysarthric speakers' intelligibility and speech characteristics in relation to computer speech recognition. *Augmentative and Alternative Communication*, *11*(3), 165–175.

Fok, R., Kaur, H., Palani, S., Mott, M. E., & Lasecki, W. S. (2018). Towards More Robust Speech Interactions for Deaf and Hard of Hearing Users. *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '18*, 57–67.

Fortune. Dragon Systems SPEECH-RECOGNITION SOFTWARE - July 6, 1998. Retrieved March 19, 2019, from http://archive.fortune.com/magazines/fortune/fortune_archive/1998/07/06/244793/index.htm

Google. Accessibility. Retrieved from https://www.google.com/accessibility/

Guzman, A. L. (2019). Voices in and of the machine: Source orientation toward mobile virtual assistants. *Computers in Human Behavior*, *90*, 343–350.

Hoy, M. B. (2018). Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. *Medical Reference Services Quarterly*, *37*(1), 81–88.

Hura, S. L. (2017). Usability Testing of Spoken Conversational Systems. *Journal of Usability Studies, 12*(4), 155-163. Retrieved from http://uxpajournal.org/wp-content/uploads/pdf/JUS_Hura_August2017.pdf

Hux, K., Rankin-Erickson, J., Manasse, N., & Lauritzen, E. (2000). Accuracy of three speech recognition systems: Case study of dysarthric speech. *Augmentative and Alternative Communication*, *16*(3), 186–196.

Juang, B. H., & Rabiner, L. R. (2004) *Automatic Speech Recognition – A Brief History of the Technology Development*. 24.

Kotler, A., & Tam, C. (2002). Effectiveness of using discrete utterance speech recognition software. *Augmentative and Alternative Communication, 18*(3), 137-146.

Kotler, A.-L., & Thomas-Stonell, N. (1997). Effects of speech training on the accuracy of speech recognition for an individual with a speech impairment. *Augmentative and Alternative Communication*, *13*(2), 71–80.

Make Your Skill Accessible to All. Retrieved from https://developer.amazon.com/docs/alexa-design/accessibility.html

Microsoft accessibility. Retrieved from https://www.microsoft.com/en-us/accessibility

Moore, M., Venkateswara, H., & Panchanathan, S. (2018). Whistle-blowing ASRs: Evaluating the Need for More Inclusive Speech Recognition Systems. *Interspeech 2018*, 466–470.

Morales, T. (2015). VoiceOver Turns 10. Retrieved from https://www.applevis.com/blog/apple-assistive-technology-ios-macos-news-opinion/voiceover-turns-10

Mulfari, D., Meoni, G., Marini, M., & Fanucci, L. (2018). Towards a Deep Learning Based ASR System for Users with Dysarthria. *Lecture Notes in Computer Science Computers Helping People with Special Needs,* 554-557.

Mullin, E. (2016) Why Siri Won't Listen to Millions of People with Disabilities. Retrieved February 7, 2019, from Scientific American website: https://www.scientificamerican.com/article/why-siri-won-t-listen-to-millions-of-people-with-disabilities/

Neerincx, M. A., Cremers, A. H. M., Kessens, J. M., van Leeuwen, D. A., & Truong, K. P. (2009). Attuning speech-enabled interfaces to user and context for inclusive design: technology, methodology and practice. *Universal Access in the Information Society*, *8*(2), 109–122.

Norman, D. (2013). *The design of everyday things*. New York: Basic Books.

Pew Research. (2017) Voice assistants used by 46% of Americans, mostly on smartphones. Retrieved March 18, 2019, from Pew Research Center website: http://www.pewresearch.org/fact-tank/2017/12/12/nearly-half-of-americans-use-digital-voice-assistants-mostly-on-their-smartphones/

Pradhan, A., Mehta, K., & Findlater, L. (2018). “Accessibility Came by Accident”: Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI ’18*, 1–13.

Resources for Developers and Publishers. Retrieved from https://www.google.com/accessibility/for-developers/

Rosen, K., & Yampolsky, S. (2000). Automatic speech recognition and a review of its functioning with dysarthric speech. *Augmentative and Alternative Communication*, *16*(1), 48–60.

Rosengren, E., Raghavendra, P., & Hunnicutt, S. (2009). How does automatic speech recognition handle dysarthric speech? *Working Papers / Lund University, Department of Linguistics and Phonetics, 43*, 112-115. Retrieved from https://journals.lub.lu.se/LWPL/article/view/2557.

Shu, C. (2019, February 26). Cedars-Sinai puts Amazon Alexa in patient rooms as part of a pilot program. Retrieved from https://techcrunch.com/2019/02/25/cedars-sinai-puts-amazon-alexa-in-patient-rooms-as-part-of-a-pilot-program/

Tecla App: Amazon Alexa. Retrieved from https://gettecla.zendesk.com/hc/en-us/articles/115004441013-Tecla-App-Amazon-Alexa-

The Smart Audio Report, Winter 2018. (2018) Retrieved from https://www.nationalpublicmedia.com/smart-audio-report/latest-report/

Voiceitt. Retrieved from http://www.voiceitt.com/

Wallace, T., & Morris, J. T. (2018). *Smart Speaker Usability by Military Service Members with mTBI and PTSD*. 13.

Yamagishi, J., Veaux, C., King, S., & Renals, S. (2012). Speech synthesis technologies for individuals with vocal disabilities: Voice banking and reconstruction. *Acoustical Science and Technology*, *33*(1), 1–5.

Young, V., & Mihailidis, A. (2010). Difficulties in Automatic Speech Recognition of Dysarthric Speakers and Implications for Speech-Based Applications Used by the Elderly: A Literature Review. *Assistive Technology*, *22*(2), 99–112.

Appendix A

Glossary

Automatic speech recognition (ASR): computerized systems that converts natural human speech into text or actions like dictation

Correct response rate: a measure of the ability of a system to correctly respond to certain commands or comments

Dysarthria: a speech disorder caused by brain damage or abnormality that produces weakness or difficulty controlling the muscles involved in speech which occurs for people with ALS, stroke, cerebral palsy, Parkinson's and more

Speech disorder: an impairment that affects, rhythm, rate, quality, pitch, loudness, and duration of speech

Speech intelligibility: the overall percentage of a person's speech that can be understood by a listener with normal hearing can understand

Speech recognition rate: a measure of the accuracy of a system to convert speech into text.

Voice assistant (VA): a speech recognition system that uses natural language input from a speaker, responds with a programmed digital voice, and carries out predetermined actions

Voice drifting: an effect of automatic speech recognition systems where the system starts to adapt to altered or fatigued speech, thereby increasing the possibility of mis-recognition when non-fatigued speech is used

Word error rate: a measure of how frequently a speech recognition system has made an error in transcription

Appendix B

Usability of Automatic Speech Recognition Systems for Individuals with Speech Disorders:
Past, Present, Future, and A Proposed Model

PART 1: Automatic Speech Recognition Systems 1990-2011

Citation	*Scope*	*Methods & Study Design*	*Data Analysis*	*Results*	*Discussion*
Doyle, P. C. (1997). Dysarthric speech : A comparison of computerized speech recognition and listener intelligibility. Journal of Rehabilitation Research and Development, 34(3), 309-316.	To identify and compare the recognition rate of dysarthric speech by IBM's VoiceType program versus human-scored speech intelligibility	6 dysarthric speakers with varying severity and 2 healthy controls were asked to produce randomized word lists over 5 sessions. Recordings from each session were used to create stimuli for a perceptual test for 10 normal hearing listeners to rate the speech intelligibility of each individual.	Recognition rates for VoiceType and average perceived speech intelligibility were calculated as percentages and plotted to compare results.	Performance of VoiceType varied by severity of dysarthria but did increase with each session for most individuals. The average speech recognition rate for dysarthric speakers was found to be 57% with the highest score achieved 80% and the lowest 35% respectively.	Speakers with severe dysarthria had lower scores of recognition but speakers with mild and moderate dysarthria were able to achieve similar scores to healthy controls for VoiceType. This indicates this system may be a useful tool for speakers with mild to moderate dysarthria.
Ferrier, L., Shane, H., Ballard, H., Carpenter, T., & Benoit, A. (1995). Dysarthric speakers' intelligibility and speech characteristics in relation to computer speech recognition. Augmentative and Alternative Communication, 11(3), 165–175.	To examine the use of the DragonDictate speech recognition system as a writing aid for dysarthric speakers, determine how speech intelligibility affects recognition rate, identify difficult to understand speech characteristics, and to track performance over several sessions.	10 individuals with spastic dysarthria and 2 control subjects were asked to read the Pledge of Allegiance for up to 8 sessions to try to achieve an 80% recognition rate. Audio recordings were judged by a speech-language pathologist to determine specific speech and voice features for each subject.	Rates of recognition were compared between the control subjects and dysarthric speakers. The total number of speech and voice features were calculated for each subject and compared to their recognition rates. Performance of DragonDictate was measured by the total number of sessions each participant needed to achieve 80% recognition rate.	All participants regardless of severity, were able to achieve at least 80% recognition rate for DragonDictate within 8 readings. Many participants were even able to achieve 90% or higher. Several common speech characteristics were identified among the participants with dysarthria.	Researchers were surprised to find that even those with severe dysarthria were able to achieve high recognition rates by this system.
Hux, K., Rankin-Erickson, J., Manasse, N., & Lauritzen, E. (2000). Accuracy of three speech recognition systems: Case study of dysarthric speech. Augmentative and Alternative Communication, 16(3), 186–196.	Compared the accuracy of Microsoft Dictation, Dragon NaturallySpeaking, and VoicePad Platinum for recognizing dysarthric speech.	1 individual with mild dysarthria and 1 control subject were asked to read 5 repetitions of certain passages to train each system. Next, participants produced 10 predetermined and 10 novel sentences.	Sentences were analyzed for speech recognition accuracy and compared across each of the 3 systems.	Dragon NaturallySpeaking was found to have about 65% accuracy while Microsoft Dictation showed about 52%, and VoicePad about 53%.	Dragon NaturallySpeaking had significantly higher recognition rates for both the dysarthric and control subjects, advocating it as a potential tool for people with speech disorders. Researchers noted that the dysarthric participant was motivated to work on this project in hopes of increasing their overall writing efficiency and independence.
Kotler, A., & Tam, C. (2002). Effectiveness of using discrete utterance speech recognition software. Augmentative and Alternative Communication, 18(3), 137-146.	Aimed to explore recognition accuracy and participant perceptions on the advantages/disadvantages of using speech recognition software.	6 participants with physical disabilities (2 with dysarthria) were observed for 1 session of dictation tasks using their own preferred speech recognition system in their home. Sessions were videotaped to calculate the recognition rates of each system and a letter from each participant was collected to gather insights on user perceptions.	From the tapes, speed of dictation and recognition rates for each systems were calculated. Responses in the letters were grouped by theme and qualitatively analyzed.	The average recognition rate for each participant was 62.4-84.4%. Participants reported being satisfied with the hands-free nature of the systems and increased productivity. They also reported several disadvantages of these systems including: unacceptable recognition rates, slowness, fatigue while using, lack of confidentiality and voice-related issues with use.	These systems mostly performed lower than recognition rates reported by other researchers. Participants saw many benefits to using these systems but also found many disadvantages primarily with the recognition rates and slowness of the system.

PART 1: Automatic Speech Recognition Systems 1990-2011

Citation	*Scope*	*Methods & Study Design*	*Data Analysis*	*Results*	*Discussion*
Kotler, A.-L., & Thomas-Stonell, N. (1997). Effects of speech training on the accuracy of speech recognition for an individual with a speech impairment. Augmentative and Alternative Communication, 13(2), 71–80.	Investigated the effects of a specific speech therapy regimen on the recognition accuracy of VoiceType (Version 1.00) for an individual with dysarthria.	1 participant with moderate dysarthria was prompted to read several passages to establish baseline speech intelligibility and to measure baseline recognition rate of VoiceType over 10 sessions. Next, the participant went through a specific speech therapy regimen and VoiceType was retested.	The performance of the system was analyzed using measures of accuracy of isolated words and words in sentences both before and after speech therapy regimen.	Participant's recognition accuracy for isolated words and words in sentences improved dramatically from 9% and 43%, respectively, to a maximum of 72% and 90% for isolated words and words in sentences.	Results indicated that specific speech therapy can lead to significantly better outcomes with speech recognition technologies.
Rosengren, E., Raghavendra, P., & Hunnicutt, S. (2009). How does automatic speech recognition handle dysarthric speech? Working Papers / Lund University, Department of Linguistics and Phonetics, 43, 112-115.	Tested the speech recognition rates of Infovox and DragonDictate systems for a severely dysarthric speaker.	1 speaker with severe dysarthria and 1 control subject trained each system with 3 utterances of around 45 words.	Performance of each system was evaluated by speech recognition score after training.	For normal speaker, recognition rates increased from 66%-75% for DragonDictate and increased from 95%-98% for Infovox. For the dysarthric speaker, recognition rates for DragonDictate increased from 31%-38% but free text speech had 82% recognition. Average recognition rate for the dysarthric speaker was 74% for Infovox.	Researchers noted that Infovox is a speaker-dependent system while DragonDictate is a speaker adaptable system. Results indicate that speaker-dependent speech recognition systems could be more effective for people with dysarthria.

PART 2: Voice Assistants 2011-Present

Citation	*Scope*	*Methods & Study Design*	*Data Analysis*	*Results*	*Discussion*
Ballati, F., Corno, F., & De Russis, L. 1 (2018). Assessing Virtual Assistant Capabilities with Italian Dysarthric Speech. Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '18, 93–101.	Investigated the extent to which people with dysarthria can use three widely used voice assistants: Siri, Google Assistant, and Cortana.	34 Italian sentences were recorded from 8 people with dysarthria, audio files from each participant were presented to each voice assistant.	Accuracy of each system was evaluated by word error rate, correct response rate, and consistency of response.	Word error rates were 24.88%, 39.39%, and 70.89% for Google Assistant, Cortana, and Siri respectively. Correct responses were 63.96%, 39.7%, and 15.81% for these systems and consistency in answers were 54.02%, 24.07%, and 60.47%.	Researchers propose that the most useful system for people with dysarthria is Google Assistant but stated that there were strengths of Siri as well.
Ballati, F., Corno, F., & Russis, L. D. 2 (2018). "Hey Siri, do you understand me?": Virtual Assistants and Dysarthria. 11.	Investigated the extent people with dysarthria can use the three most common virtual assistants: Siri, Google Assistant, and Amazon Alexa.	5 sentences were created from the TORGO dataset pronounced by 7 speakers.	Performance of each system was measured by correct response rate of each system and consistency of answer.	Siri had a recognition rate of 58.82%, Google Assistant 64.7%, and Alexa 41.17%.	Results were consistent with previous findings for automatic speech recognition for people with dysarthria.
Moore, M., Venkateswara, H., & Panchanathan, S. (2018). Whistle-blowing ASRs: Evaluating the Need for More Inclusive Speech Recognition Systems. Interspeech 2018, 466–470.	Tested Carnegie Mellon University's Sphinx Open Source Recognition, and Google Speech Recognition ability to recognize dysarthric speech.	Both systems were tested for transcription ability using the TORGO and UASPEECH datasets.	Word error rates were calculated for each system and compared to the word error rate judged by normal hearing human listeners. From this recognition rate was calculated.	As expected, speakers with lower speech intelligibility had lower recognition rates for both systems. It was found that Sphinx had a 7.2% average recognition rate and Google 29.64%.	Analysis demonstrated that automatic speech recognition systems do not provide robust speech recognition to individuals with voices that fall outside the range of 'normal' voices.
Pradhan, A., Mehta, K., & Findlater, L. (2018). "Accessibility Came by Accident": Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18, 1–13.	Examined the accessibility of Amazon's voice assistant series for users with disabilities using a content analysis of 346 Amazon Echo reviews.	Methods and study design: Researchers located Amazon reviews for Amazon Echo and Echo Dot from verified users and extracted reviews that contained defined keywords related to cognitive, sensory, or physical abilities.	Reviews were coded along 26 defined dimensions by two researchers and were qualitatively analyzed.	346 reviews were identified relating to users with disabilities. Of these 13.6% mentioned speech impairment. The largest accessibility concern with these systems was for people with speech impairments. Surprisingly, many of the reviews of people with speech impairments were positive.	Despite these systems being highly accessible, challenges arose for people with speech impairments. Regardless, many reviews were positive and demonstrate that this group wants to use this technology.

Usability of Automatic Speech Recognition Systems for Individuals with Speech Disorders:
Past, Present, Future, and A Proposed Model

Madeline Jefferson
University of Minnesota, Spring 2019

Introduction

•People are using voice assistants (VAs) such as Siri and Alexa more than ever before.
•VAs provide convenience, novelty, and unique solutions for the medical industry. But users with speech disorders may be left out of the conversation.
•Historically, usability testing for ASRs, the precursor to VAs, has consistently shown they are not easy to use for people with speech disorders like dysarthria.
•This investigation sought to perform a two stage literature review on the existing research on the usability and user perceptions of commercially available ASRs & VAs for people with dysarthria, a common speech disorder.
•Doing so provides historical perspectives on this topic, explores how this issue is being addressed today, and helps us understand how technical communicators and usability professionals can best address the needs of this audience.

Glossary

Automatic speech recognition (ASR): computerized system that converts human speech into text or actions like dictation

Correct response rate: measure of the ability of a system to appropriately respond to certain commands or comments

Dysarthria: speech disorder caused by brain damage or abnormality that produces weakness or difficulty controlling the muscles involved in speech

Speech recognition rate: measure of the accuracy of a system in converting speech into text

Voice assistant (VA): speech recognition system that responds with a programmed digital voice and carries out predetermined action

Word error rate: measure of how frequently a speech recognition system has made an error in transcription

Methods

•The literature review was performed in two parts on the usability of commercially available ASRs from 1990 to 2011 (Part 1) and on the commercially available VAs from 2011 to Present (Part 2) for people with dysarthria.
•After searching several databases using two sets of predefined keywords and evaluating results with explicit exclusion criteria, 12 studies were chosen for analysis.

Results

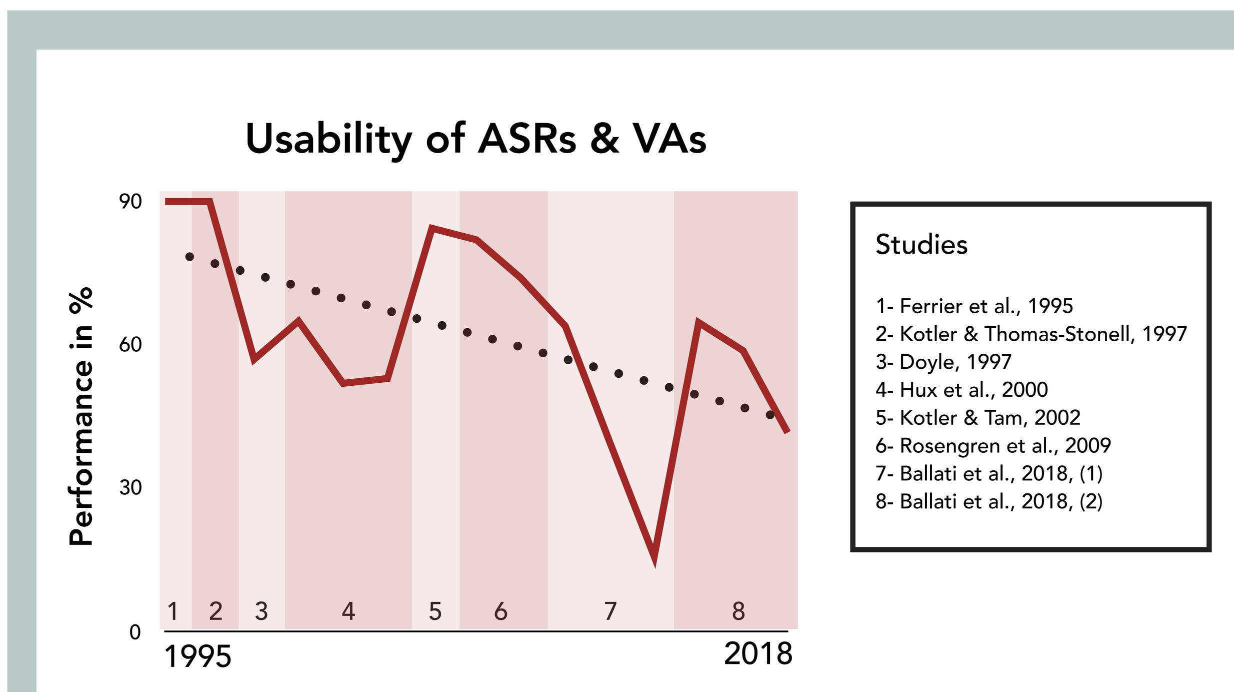
•Part I 1990-2011: many of the studies on ASRs found surprisingly high speech recognition rates for individuals with a variety of degrees of dysarthria: 62-80%.
•Part II 2011-Present: The majority of commercially available VAs performed with a 25-60% correct response rate.
•User Perceptions: Both studies that explored user perceptions and attitudes found individuals with dysarthria and speech disorders were frustrated with these systems and found them time-consuming. However, users saw value in the hands free nature of this technology and wanted to use it regardless.

Why are these companies making voice-based systems unusable for the people who could benefit from them the most?

amazon
Microsoft
Google

Discussion

•Performance of ASRs pre-2011 was significantly higher than that of the current off-the-shelf voice assistants indicating that the usability of these systems for people with speech disorders has regressed.
•People with speech disorders want to use these systems despite their reduced usability for them.
•The mission statements and developer guidelines of the tech companies designing the most popular VAs fail to mention how they are addressing this issue.
•This investigation led to a proposed model intended to guide technical communication and usability professionals on how to make VAs usable for individuals with speech disorders.



The data tells the story of a technology that arose as a usable tool for people with dysarthria but now functions as a one-size-fits-most tool, systematically excluding people with speech disorders.

Proposed Model

Train with Diverse Speech
Datasets used to train VAs need to be more inclusive of unique voices to ensure that the system is more likely to respond to the speech characteristics of users with speech disorders.

Use Multiple Modalities
Including multiple modalities, such as a camera for facial recognition, would increase the likelihood of the system understanding what the user is trying to say.

Optimize the Signal
We can reduce the burden of usability for these individuals by training these systems to respond better to atypical speech behaviors such as repetitions, pauses and a slower speaking rate.

Make Sytems Speaker-depedent
Calibrating systems to specific speakers, while requiring more set up, would allow VAs to understand how each user pronounces every word and sound with their own voice.

Use Machine Learning
There are specific techniques that can be used to account for voices that change over time and with fatigue. This would be particularly effective for users with neurodegenerative disorders like ALS.

Create Speech Disorder Friendly Content
Technical communicators and developers can reduce the number of words needed for commands/skills and avoid specific sounds known to be difficult to pronounce for people with speech disorders.

For further questions contact: jeffe150@umn.edu