Deaf, Hard of Hearing, and Hearing Perspectives on using Automatic Speech Recognition in Conversation

Abraham Glasser, Kesavan Kushalnagar, and Raja Kushalnagar
Rochester Institute of Technology
Rochester, NY 14623
{atg2036, krk4565}@rit.edu, raja.kushalnagar@gallaudet.edu

ABSTRACT

Many personal devices have transitioned from visual-controlled interfaces to speech-controlled interfaces to reduce costs and interactive friction, supported by the rapid growth in capabilities of speech-controlled interfaces, e.g., Amazon Echo or Apple's Siri. A consequence is that people who are deaf or hard of hearing (DHH) may be unable to use these speech-controlled devices. We show that deaf speech has a high error rate compared to hearing speech, in commercial speech-controlled interfaces. Deaf speech had approximately a 78% word error rate (WER) compared to a hearing speech 18% WER. Our findings show that current speech-controlled interfaces are not usable by DHH people. Based on our findings, significant advances in speech recognition software or alternative approaches will be needed for deaf use of speech-controlled interfaces. We show that current speech-controlled interfaces are not usable by DHH people.

CCS Concepts

•Human-centered computing → Accessibility; Accessibility design and evaluation methods;

Keywords

Automatic Speech Recognition, Deaf Speech, Deaf, Hearing

1. INTRODUCTION

About 30 million people in the United States have bilateral hearing loss [1], and about 1 million are functionally deaf [2]. Speech production quality is correlated with hearing loss [3], which can lead to both speaking and listening difficulties. Around half a million people in United States communicate visually through American Sign Language (ASL), and use it as their primary means of communication [4]. Deaf or hard of hearing (DHH) people usually cannot understand speech unaided, and usually depend on additional support such as hearing aids or speech-to-text technology, compared with their hearing (H) peers. Simple low-technology aids such as using paper and pen to write back and forth or to text back and forth can work, but are about 3-4 times slower than spoken or signed communication, and is not effective for sustained communication.

ASSETS '17, Oct. 29–Nov. 1, 2017, Baltimore, MD, USA.

ISBN 978-1-4503-4926-0/17/10. . . $15.00.
DOI: https://doi.org/10.1145/3132525.3134781

1.1 Automatic Speech Recognition

Many personal devices offer aural interfaces (e.g., phones) that use Automatic Speech Recognition (ASR) and Text-To-Speech (TTS). Aural interfaces are being rapidly adopted as they have become reasonably accurate and easy-to-use, through significant advancements in speech recognition, machine learning and context sensing services. With increasing interest in IoT (Internet of Things) devices, voice-controlled personal assistants have also become popular as standalone computing devices in home and office environment, as they are cheap, convenient and reasonably accurate.

1.2 ASR Barriers

DHH users cannot easily use these aural interfaces because the ASR services used in these interfaces are inaccurate at recognizing deaf people's speech, which have wide variation and disfluencies, even for short commands and requests [5]. The acoustic and linguistic characteristics of speech associated with DHH people is different from non-DHH people, and usually varies dramatically as a function of hearing loss and onset [3, 6]. As a consequence, Automatic Speech Recognition (ASR) systems trained on speech from non-DHH people perform poorly for recognizing deaf speech [7]. In particular, even if the deaf person had highly intelligible speech, commercial ASR services could not recognize many spoken words, and participants were dissatisfied with the service [8].

1.3 Conversational Barriers

DHH and H people face diverse challenges in spoken language communication with each other in conversational settings, especially in multiple-talker settings such as in classrooms and workplaces. In contrast with H listeners, DHH listeners have manage competing tasks such as shifting attention between multiple visuals. They have to connect incomplete segments together, while searching for cues to know where to pay attention. As a result, even when provided with accurate real-time text through captioners, they receive only 50%-80% of the information, compared to 84%-95% for H peers [9]. Similarly, HH participants have to make sense of reduced speech information through their hearing aids. Both groups need flexible accessible technology solutions for upward mobility [10].

2. PERSPECTIVES

This experience report describes the accessibility challenges by two deaf, one hard of hearing and two hearing participants, including the authors, in using Automatic Speech Recognition (ASR) applications on personal devices for commands and group conversation. Deaf, hard of hearing and hearing speakers and listeners have different challenges and accessibility needs in mixed group conversation in most settings, including academic and workplace settings. We discuss how current ASR applications enhance access by deaf and hard-of-hearing individuals. We also examine how ASR applications enhance communication exchanges between deaf or hard-of-hearing persons and hearing persons in the classroom or workplace.

2.1 Deaf Participants

The deaf participants reported that their challenge was in accessing and following spoken information and in conveying information efficiently and quickly to others in group settings. They preferred to use ASR for one-to-one conversation in quiet settings, so that ASR could accurately display and read the words of their communication partners. Sometimes they also use ASR to speak and convey information quickly to their communication partners especially if their speech clarity is sufficiently close enough to that of their hearing peers.

2.2 Hard of Hearing Participants

The hard of hearing participant usually did not have conversational challenges in quiet or one-to-one settings. Instead, the participant often had difficulties in multi-speaker or noisy settings, because the audio signal was degraded, or with multiple information sources, and/or talkers with dialects or accents [11].

Even with the latest hearing aids that incorporate noise reduction algorithms that are capable of improving listening-alone performance, the participant was not able to make up for the adverse effects of having to concentrate on following speech, and or dealing with competing tasks such as taking notes or shifting attention between multiple speakers or visuals, which is common for hard of hearing listeners [12].

2.3 Hearing Participants

The hearing participants did not have difficulty in speaking or listening to other hearing peers in most settings, but had difficulties in conversing with deaf or hard of hearing speakers, especially if they did not have prior experience with the deaf or hard of hearing person. Hearing listeners use similar cognitive-perceptual processing adaptations to better understand speech under adverse listening conditions, either noise (environmental degradation) or dysarthric speech (source degradation) [13].

2.4 Deaf-Hearing Communication

Deaf or hard of hearing speakers usually have wide variance in speech production. Figure 1 shows the distribution of speech samples from about 650 deaf people as rated by speech pathologists [14]. Each of these samples and ratings were based on an assessment of intelligibility and comprehensibility, using a set of sentences called "Clarke Sentences"[1] with ratings from 0 to 50.

[1]https://www.ntid.rit.edu/slpros/assessment/intel/sentence

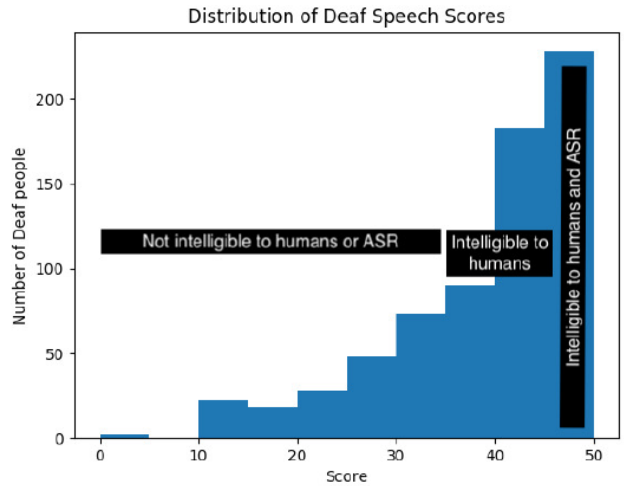

Figure 1: Distribution of Deaf Speech scores.

Hearing listeners usually cannot adapt and understand DHH speakers with ratings of between 0 to 30. If they have a lot of experience listening to deaf speech, they can adapt their cognitive processing strategies to follow DHH speakers with ratings between 30 to 45. If they do not have prior experience, they can follow DHH speakers with ratings from 45 to 50 with some effort. So, hearing speakers with little prior experience are able to understand fewer than half of DHH speakers on average.

2.5 ASR Services

ASR services are rapidly improving due to large investments in aural interfaces for use in wearables, cars, robotics, and machines. The incorporation of large datasets with millions of speakers has led to higher accuracy with a wider range of speech patterns, such as Microsoft at 6% in late 2016 [15] and Google at 5% in early 2017[16].

However, ASR services are trained with large speech sets under good audio conditions and adapt poorly to adverse audio conditions, either noise or speech [17]. While new algorithms have shown promise in handling the dysarthric speech variation [18], they need large datasets, which can be difficult to get. An ASR evaluation with the Clarke Sentence samples drawn from DHH speakers rated 5.0, yielded a Word Error Rate (WER) of about 53% [14], which is generally too high. Other ASR problems include lag, jitter, and acquisition factors such as fidelity, ambient noise and microphone quality. ASR services still had word error rates of 20-25% for lectures [11].

2.6 Accuracy

The text accuracy for real-time speech-to-text needs to be sufficiently high to be useful. Studies have shown that DHH people will use real-time speech-to-text, such as ASR, if its accuracy is least 85% [19] to 90% [20] or higher.

2.7 Lag and Jitter

The lag time for ASR needs to be short enough to be usable. Lag time for ASR becomes worse as the amount of data to be analyzed increases and causes processing delays. Similarly variance in processing time significantly bothers

Figure 2: Logos of various ASR platforms used

users. They cannot effectively participate in discussions or dialogues if lag time is more than 5 seconds [19], or if variance is more than 2-3 seconds [11].

3. METHODS

To investigate the capabilities of current ASR applications, from Fall 2016 through Summer 2017, five participants used one or more of seven ASR applications on their personal devices in everyday, real-world settings for conversations.

3.1 Users

One of the deaf participants did not use spoken conversation with hearing peers and typed on the phone and used ASR for conversing with non-signing peers. The other deaf participant used spoken conversation with hearing peers in face-to-face conversations, and used ASR in group conversations.

The hard of hearing participant used spoken conversation with hearing peers in face-to-face conversation, and used ASR in group conversations.

The hearing participant used spoken conversation with hearing peers in both face-to-face and group conversations, and used either ASL or ASR with the hard of hearing and deaf peers.

3.2 Contexts

The purpose of using the ASR applications was to assess the usability of the applications in face-to-face spoken language interactions by providing a visible text representation of speech in the following contexts:

- Classroom communication
- Job Interviews
- Conversation
- Speech production practice

3.3 Applications

The participants used and evaluated the following ASR apps: DEAFCOM, Dragon Dictation, Siri, Virtual Voice, Ava, Google Assistant, and Amazon Alexa, as shown in Figure 2. These apps were chosen because they were available for free and had been rated at least 3.5 out of 5 for user satisfaction in the store.

1. DEAFCOM by askjerry Communications is aimed at DHH to use for conversation, with a simple interface that is quick to start. Its description is: ".... For a non-hearing or hard-of-hearing person, the application will allow faster communication with deaf persons. For deaf users, the software can assist in faster communication and may also be used as a useful tool when practicing your speech".[2]

2. Dragon Dictation by Nuance Communications is for all users as a voice recognition system: " ... an easy-to-use voice recognition application powered by Dragon ...."[3]

3. Siri by Apple, is for all users as an aural interface assistant: "Talk to Siri as you would to a friend and it can help you get things done, like sending messages, placing calls, or making dinner reservations ...."[4]

4. Virtual Voice by Gareth Hannaway Communications is aimed at DHH to use for conversation, with a simple interface that is easy to read: "It is designed to use the text to speech (TTS) and the speech recognition features of your Android device. It was created with deaf and/or mute people in mind, so they can communicate with others without the need for sign language or lip reading".[5]

5. Ava by Ava, is aimed at DHH to use for conversation in pairs or groups: "Ava shows you who says what. Ava shows you what people say, in less than a second. Easy communication is only a tap away."[6]

6. The Google Assistant by Google, is for all users as an aural interface assistant: "Meet your Google Assistant. Ask it questions. Tell it to do things. It's your own personal Google, always ready to help."[7]

7. Alexa by Amazon, is for all users as an aural interface assistant: "Using Alexa is as simple as asking a question.The more you talk to Alexa, the more it adapts to your speech patterns, vocabulary, and personal preferences."[8]

3.4 Evaluation

The participants downloaded the apps to their personal iPhone or Android device (or directly used the device), and evaluated their use in the contexts listed above.

[2]play.google.com/store/apps/details?id=defcom.v1
[3]www.dragonmobileapps.com/android/
[4]www.apple.com/ios/siri/
[5]play.google.com/store/apps/details?id=appinventor.ai_Gareth_Hannaway_420.VirtualVoice
[6]https://www.ava.me/
[7]https://assistant.google.com/
[8]https://www.amazon.com/Amazon-Echo-And-Alexa-Devices/b?ie=UTF8&node=9818047011

All participants reported that when they used the apps under ideal circumstances, the captions spoken by their hearing peers were accurate, with minimal lag and jitter. The ideal circumstance was to use the app with excellent WiFi coverage, in quiet one-to-one settings with hearing peers with American accents when discussing general topics. The apps did well in all contexts, when used for five minutes or less, regardless of whether it was used in a classroom, informal conversation, job interviews or speech production practice in which the text displayed by an app was used as an indicator of the intelligibility.

3.5 Duration

After continuous conversation for more than five minutes, five of the seven apps (interactive assistants and apps for the deaf, but not ASR service apps) showed significant time lag in quiet settings, especially in classroom settings, which tend to have multiple speakers who speak fast. The hearing and hard of hearing participants reported that on some occasions the time lag resulted in dropped words, while the deaf participants did not report this phenomenon.

3.6 Noise

All apps exhibited significant lag more quickly (within a minute) and jitter (variance) when there was background noise, which happened often in all contexts. They all inserted random text as well. The lag occurred even when the background noise was not noticeable to hearing peers in the room.

Accuracy was also disappointing when there was any level of noise in the environment. The combination of inaccurate transcription and time lag was very frustrating for the deaf and hard of hearing participants, and they reported that there was much work yet to be done to make these apps useful enough to put away pencil and paper.

3.7 Multiple speakers

All apps had significant lag more quickly (within a minute) and jitter (variance) when there were multiple speakers, in all contexts. The lag occurred even when the background noise was not noticeable to hearing peers in the room.

For most applications, the deaf users could not tell who was speaking and often became confused and frustrated. On the other hand, hearing peers who had concerns about communicating with deaf or hard of hearing individuals felt more at ease when they were able to use the apps, even in noisy or multi-speaker settings.

3.8 Hearing Accents

All apps also had more lag, jitter and were less accurate when the speech had variance from standard American speech patterns when the speaker had an accent from another country. In addition to being less accurate, the apps often inserted random text.

3.9 Deaf Speech/Accent

Most apps did not recognize the different prosody, pitch, and articulations of deaf speakers, even for the apps that were aimed at use by deaf or hard of hearing users. These DHH specific apps had less lag time, but still had high error rates. For purposes of facilitating speech intelligibility, most of the DHH evaluators, regardless of their mode of communication, did not find the speech recognition apps to be usable. The DHH evaluators were quite disappointed to find that they uniformly failed to recognize their speech and they tended to attribute this failure to differences in their own articulation, pitch, and prosody. The alternative of switching to a text-to-speech or text-to-text function resulted in significant slowing in conversational interactions.

4. LESSONS LEARNED

When using ASR technology, deaf, hard of hearing and hearing people report different ways of interacting with the technology. The following lessons were learned from daily experience with the technology. In most situations, the settings were not ideal, and the level of accuracy and degree of latency characteristic of the apps were not adequate to enhance speech reception in face-to-face interaction. We experienced a disruptive display of text that did not match in time what the others were saying. More needs to be done so that the text is synchronous with the speech. When the authors ignored the speech or did not speechread, it was useful to have access merely to the overall context of the message via keywords that the apps could display.

In summary, regardless of context, the main factors for app usability were:

- Noise (e.g., music)
- Speech produced with an accent
- Multi-talker speech or side conversation
- Disfluent speech or speakers with emotion.

The shared use of ASR apps by both deaf and hearing people drew them into a collective learning experience:

4.1 Deaf users

> We are not the only ones with problems understanding .... That we're not the only ones. There's non-disabled people out there who are having the same problems ... you feel equal.

> The idea today is that everyone should make a small effort to make the conversation work. Of course a minimal effort, but the burden should not be on the deaf person alone. Everyone works together.

> The app is difficult to use. I prefer to use paper and pen as it is more reliable and easier to use.

4.2 Hard of hearing users

> It's telling me that at least I'm not the only person that might have a problem understanding. Like, I know that sometimes when you've got a disability you feel like you're the only one ... I just don't want it to benefit us. I'd like to see it work for everybody

> I cannot use Alexa at my parents' house, as I cannot understand its responses after I give it a command.

4.3 Hearing users

> I am very satisfied with using ASR apps for looking up simple queries on the web. I did not realize that the apps are not accurate or easy to use for conversational use with deaf friends.

5. CONCLUSIONS

ASR has been constantly updated and improved for over 50 years. However, ASR has been and continues to be focused on hearing speakers who have low variance and in their speech. It is difficult for deaf and hard of hearing individuals, even those who use voice on a regular basis, to be fully comfortable with ASR. They cannot dictate to ASR services reliably, because there is a big variance their speech, even if their speech is understandable by their hearing peers. Hearing peers use additional information such as experience with common deaf speech patterns, visual cues to accurately understand the DHH speaker. So, even if the deaf person can be understood by hearing peers, all ASR services currently tested do not reliably provide accurate or usable transcripts.

5.1 User Interface

DHH individuals report that the hearing peer's patience and attitude matters, especially when the hearing peer has never used ASR before. The DHH individuals note that hearing peers often get frustrated with the user interface design of the ASR app itself, since these interfaces are not intuitive to use for conversational use. The DHH users said that it would be ideal if ASR systems could tell the user if repetition of a specific word was needed rather than the whole word. Using a system like this might increase the interaction of users with the device.

5.2 Visual Interface Feedback

A big barrier to interaction with ASR devices is that DHH individuals do not have access to the verbal output from these devices after speaking commands to it, i.e., verbal inputs. Many interfaces are voice only, and are inaccessible to many DHH individuals. Some ASR devices are able are able to display verbal responses as text on the screens of paired personal phones. However, many new devices on the market do not have any visual interface and and do not connect with personal phones.

A couple of personal phones with ASR services have recently added text input capability in addition to voice input capability, which makes these services more accessible for DHH individuals. But because text interfaces are slower than speech interfaces, these text interfaces are not as appealing as voice interfaces for using these devices as a personal assistant. DHH individuals typically cannot monitor their own speech inflections and volume, and it is important to provide users good feedback about good placement of the phone and associated speech volume.

Although some DHH individuals are able to use ASR systems such for transactional use, almost all DHH individuals are uncomfortable using these systems for sustained conversational use, as the systems have higher than tolerable error rates, especially in less than perfect settings. Whenever there are errors, the errors are time consuming to fix and the presented text cannot be edited for accuracy or clarification.

Putting aside the problems facing deaf people using their own voices with ASR, there are still user interface accessibility issues. For example, if it was practical for ASR technology to be used in conversations between multiple deaf and hearing individuals, the ASR interface should be quick to open, and display contextual cues, such as speaker identification.

5.3 Algorithms

Although ASR services show significant improvement in laboratory settings [16, 17, 21], DHH and hearing users experience significant performance issues, especially in the critical settings of the classroom and workplace. New technological advances and enhanced engineering are needed to bring the errors down much further in order to make speech recognition useful under difficult yet practical and realistic conditions [22], particularly to control noise and side-talk interference, perhaps with better noise canceling algorithms, more advanced microphone array techniques, and through use of a lapel mic, Bluetooth streaming, and/or Wi-Fi Direct, and to ultimately convince deaf and hard-of-hearing persons that speech recognition technologies are better than pencil and paper when trying to communicate to a person with typical hearing when it counts in the classroom or on the job.

5.4 Speech Quality

In addition to noisy backgrounds, many speakers, especially those who are deaf, have low volume, and directional microphones can make a significant difference. It would be helpful to include external microphone support for microphones that can be either plugged in or connected with Bluetooth. This could be a lapel type of mic that is clipped on the person, or a more central hardware piece with multiple microphones. These could indeed do both directional beam forming or omni directional capturing. Most applications need fairly loud speech samples for optimal accuracy, that is, the speech signal has to be much larger than the background noise.

5.5 App Support

People have frustrations with lack of Internet connections and battery and space usage, so it should be an objective to make ASR apps efficient. People who have little or no experience using ASR technology may not know how to interact with the app. Thus, clear, intuitive user interfaces need to be added to the ASR apps so that the conversation will have better flow and comfort. Apps are typically limited and have a certain amount they will transcribe in a time period. It has been observed that some apps require purchases for more usage. This is disadvantageous because Deaf people should not have to work harder and give up more to have equal access to information.

6. RECOMMENDATIONS

In order to improve ASR service support for conversational use between deaf, hard of hearing and hearing users, it is critical to include their perspectives and experiences in their ASR app use. Evaluations and surveys need to recruit people with different backgrounds and use. This way, the developers and researchers will be able to work with accessibility in mind and have their products benefit a wide spectrum of people.

6.1 Aural Interface Access

The increased popularity of wearables and personal devices with aural interfaces is likely to pose an significant barrier for use of aural interfaces. Most aural interfaces' ASR services are inaccurate at recognizing deaf people's speech, because their speech has wide variation and disfluencies, even for short commands and requests, let alone long speeches. They cannot easily use or interact with the aural interfaces in today's wearables or devices. Deaf users prefer to use interfaces through which they can interact with high accuracy, low variance and low lag time. Although communicating through typing and writing has high accuracy and low variance, they are slow and introduce a lot of lag time. For deaf individuals with low intelligibility ratings, it is unlikely ASR services will be able to recognize their speech, due to the high variance in their speech.

6.2 Visual Interface Access

Deaf signers, like hearing speakers, have high accuracy, low variance and low lag time in communication through sign language. Deaf signers would prefer and benefit from a visual interface that can recognize and produce sign language, and has high accuracy, low variance and low lag time. Similarly, hard of hearing people regardless of whether they pick a aural, visual, aural-visual or aural-visual-tactile interface, would prefer that it have high accuracy, low variance and low lag time.

Writing or typing is used as a last resort by both hearing and deaf or hard of hearing individuals in conversational use. Although it is understood by both parties, and is accurate, it is far slower than speaking or signing. The text representation carries less emotion and inflections, when compared to speech or sign, even the text representation is supplemented with emojis.

7. REFERENCES

[1] Frank R. Lin, John K. Niparko, and Luigi Ferrucci. Hearing Loss Prevalence in the United States. *Archives of Internal Medicine*, 171(20):1851, nov 2011.

[2] B B Blanchfield, J J Feldman, J L Dunbar, and E N Gardner. The severely to profoundly hearing-impaired population in the United States: prevalence estimates and demographics. *Journal of the American Academy of Audiology*, 12(4):183–9, apr 2001.

[3] Sven L. Mattys, Matthew H. Davis, Ann R. Bradlow, and Sophie K. Scott. Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, 27(7-8):953–978, sep 2012.

[4] Ross E. Mitchell, Travas A. Young, Bellamie Bachleda, and Michael A. Karchmer. How Many People Use ASL in the United States? Why Estimates Need Updating. *Sign Language Studies*, 6(3):306–335, 2006.

[5] C Jeyalakshmi, V Krishnamurthi, and A Revathy. Deaf speech assessment using digital processing techniques. *Signal & Image Processing: An International Journal (SIPIJ)*, 1(1):14–25, 2010.

[6] Leo De Raeve. A Longitudinal Study on Auditory Perception and Speech Intelligibility in Deaf Children Implanted Younger Than 18 Months in Comparison to Those Implanted at Later Ages. *Otology & Neurotology*, 31(8):1261–1267, oct 2010.

[7] Kinfe Tadesse Mengistu and Frank Rudzicz. Comparing Humans and Automatic Speech Recognition Systems in Recognizing Dysarthric Speech. pages 291–300. 2011.

[8] Linda G. Gottermeier and Raja S. Kushalnagar. User Evaluation of Automatic Speech Recognition Systems for Deaf-Hearing Interactions at School and Work. *Audiology Today*, 28(2):20–34, 2016.

[9] Marc Marschark, Patricia Sapere, Carol Convertino, and Jeff Pelz. Learning via direct and mediated instruction by deaf students. *Journal of deaf studies and deaf education*, 13(4):546–561, 2008.

[10] Marc Marschark, Venetta Lampropoulou, and Emmanouil K Skordilis. *Diversity in Deaf Education.* Oxford University Press, 2016.

[11] Raja. S. Kushalnagar, Walter S. Lasecki, and Jeffrey P. Bigham. Accessibility Evaluation of Classroom Captions. *ACM Transactions on Accessible Computing*, 5(3):1–24, 2014.

[12] Karen L. Anderson. Access is the issue, not hearing loss: New policy clarification requires schools to ensure effective communication access. *SIG 9 Perspectives on Hearing and Hearing Disorders in Childhood*, 25(1):24–36, 2015.

[13] Stephanie A. Borrie, Melissa Baese-Berk, Kristin Van Engen, and Tessa Bent. A relationship between processing speech in noise and dysarthric speech. *The Journal of the Acoustical Society of America*, 141(6):4660–4667, jun 2017.

[14] Abraham Glasser, Kesavan Kushalnagar, and Raja Kushalnagar. Feasibility of Using Automatic Speech Recognition with Voices of Deaf and Hard-of-Hearing Individuals. In *The 19th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 334–336, 2017.

[15] W Xiong, J Droppo, X Huang, F Seide, M. Seltzer, Andreas Stolcke, D Yu, and G Zweig. The Microsoft 2016 conversational speech recognition system. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5255–5259. IEEE, mar 2017.

[16] John Shinal. Making sense of Google CEO Sundar Pichai's plan to move every direction at once. *CNBC*, May 2017.

[17] Ming Tu, Alan Wisler, Visar Berisha, and Julie M. Liss. The relationship between perceptual disturbances in dysarthric speech and automatic speech recognition performance. *The Journal of the Acoustical Society of America*, 140(5):EL416–EL422, Nov 2016.

[18] Myungjong Kim, Younggwan Kim, Joohong Yoo, Jun Wang, and Hoirin Kim. Regularized Speaker Adaptation of KL-HMM for Dysarthric Speech Recognition. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, pages 1–1, 2017.

[19] Mike Wald. Using automatic speech recognition to enhance education for all students: Turning a vision into reality. In *Frontiers in Education, 2005. FIE'05. Proceedings 35th Annual Conference*, pages S3G–S3G. IEEE, 2005.

[20] Richard Kheir and Thomas Way. Inclusion of deaf students in computer science classes using real-time speech transcription. *ACM Sigcse Bulletin*, 39(3):261–265, 2007.

[21] Wayne Xiong, Lingfeng Wu, Fil Alleva, Jasha Droppo, Xuedong Huang, and Andreas Stolcke. The microsoft 2017 conversational speech recognition system. Technical report, August 2017.

[22] Dong Yu and Li Deng. *Automatic speech recognition: A deep learning approach*. Springer, 2014.