Assistive Technology Outcomes and Benefits
Volume 16 Issue 2, Summer 2022, pp. 45-55
 ISSN 1938-7261
Available online: www.atia.org/atob

Voices from Industry

Access for Deaf and Hard of Hearing Individuals in Informational and Educational Remote Sessions

Sheryl Ballenger, Ph.D., CPACC

Center for Inclusive Design and Innovation
College of Design
Georgia Institute of Technology

Corresponding Author
Sheryl Ballenger, Ph.D.
Center for Inclusive Design and Innovation
College of Design, Georgia Institute of Technology
512 Means Street
Suite 250
Atlanta, GA 30318
Phone: (404) 894-8679
Email: sheryl.ballenger@design.gatech.edu

Author Note

Regarding person-first language: Members of the Deaf community prefer to be referred to by their identity as a Deaf person rather than a person who is deaf. This article uses the phrase "Deaf and hard of hearing individuals."

ABSTRACT

Opportunities to present to remote audiences require access for people with disabilities. The COVID-19 pandemic, with an imperative of social distancing, provided access challenges. Innovative tools, such as artificial intelligence (AI), as used in Automatic Speech Recognition (ASR), became available in many applications. Some community information and higher education programs considered supplying access through ASR text produced by AI software tools. This article's contribution to the field is a comparative

analysis of some ASR software used as speech-to-text accommodations for Deaf and hard of hearing individuals in informational and educational settings. Some nuances of ASR and human Speech-to-Text-Services (STTS) practices are included. The concept of use cases for low-stakes settings and high-stakes settings are introduced. This article also provides a framework for future studies of the efficacy of ASR software used as an accommodation and best practices for using ASR software in informational and educational remote sessions.

Keywords: speech-to-text, artificial intelligence (AI), automatic speech recognition (ASR), auto-generated captions, real-time captioning, accessibility

ACCESS FOR DEAF AND HARD OF HEARING INDIVIDUALS IN INFORMATION AND EDUCATIONAL REMOTE SESSIONS

People who are hard of hearing have a hearing loss ranging from mild to severe. Both volume and clarity are necessary for understanding speech. A person may hear that speech is occurring, but discerning the actual words requires the ability to hear the phonemes of each sound. The term “Deaf” refers to individuals who are culturally Deaf and individuals who experience a profound hearing loss with little or no hearing (World Health Organization, 2020). Members of the Deaf community prefer to be referred to by their identity as a Deaf person rather than a person who is deaf. This article uses the phrase, Deaf and hard of hearing individuals.

Communication access for Deaf and hard of hearing individuals is essential in every aspect of their lives. The Rehabilitation Act of 1973, Section 508, obligates entities sharing informational or educational information remotely to provide access to the audio for Deaf and hard of hearing individuals. When considering remote formats such as webinars, remote meetings or training, public health and safety information, or educational sessions, organizers must be prepared to address the needs of a broad audience with varied abilities (Section508, n.d.). In order to provide access for Deaf and hard of hearing participants, organizers must ensure that audio is visible in remote sessions. Audio must be visible for Deaf and hard of hearing individuals in remote sessions (Alsalamah, 2020).

Before and during the COVID-19 pandemic, the issue of communication disparities for Deaf and hard of hearing individuals was well documented. (Barnett et al., 2011; Hoglind, 2018; National Association for the Deaf, 2021; National Deaf Center on Postsecondary Outcomes, 2020). The novel COVID-19 health information was not accessible to Deaf and hard of hearing individuals. Compounding the existing situation, the usual in-person channels of communication had suddenly disappeared. College classes and school programs moved online, workplaces closed or moved online, and in-person community events and religious services were canceled. Without a contingency plan for continued communication access, the immediate move to online or remote meant no access for Deaf and hard of hearing individuals (Hodges et al., 2020).

In the United States, in March 2020, during the early stages of the COVID-19 pandemic, meetings,

training, college courses, classes, small groups, etc., transitioned to being hosted on remote collaboration tools (Hodges et al., 2020). Faced with the imperative of social distancing, entities providing informational or educational sessions looked to the most expedient audio access option that was available, Automatic Speech Recognition (ASR; Ellcessor, 2020).

A few remote collaboration tools began rolling out ASR as a contrivance. As educators struggled with emergency remote sessions, so did some presenters of community information regarding health guidance. The Center for Inclusive Design and Innovation (CIDI) at Georgia Institute of Technology, a service and research supercenter, informs about best practices in creating accessible environments. When face-to-face interactions shifted to remote during the COVID-19 pandemic, we noticed the abrupt change from human-provided Speech-to-Text-Services (STTS) to either the sudden cancellation of provided access or to the use of ASR-provided access.

Background

The prospect of machines simulating human behaviors has been a dream of humans for hundreds of years (Ekbia & Nardi, 2014). ASR was first related to pattern recognition. The ASR and pattern recognition work are based on computers' power to recognize a pattern in speech and transform it into pertinent information (O'Shaughnessy, 2008). Reporting on the history of ASR, O'Shaughnessy (2008) noted that Bell Labs demonstrated small-vocabulary recognition for digits spoken over the telephone in 1952. Medical reporting and legal dictation implemented ASR in the 1990s, as well as telephone services (Bogdan, 2019; O'Shaughnessy, 2008). Successful ASR started to understand large vocabularies in uncontrolled environments in the 2000s (Bogdan, 2019). YouTube began providing ASR within its platform in 2009 (Alberti & Bacchiani, 2009). Parton's (2016) study showed YouTube's ASR produced 7.7 phrases that were unintelligible or that altered the meaning of the message every minute. Although Parton's (2016) study used videos, the ASR was the best available and is relevant data.

PERSONAL STATEMENT

As an experienced professional working with Deaf and hard of hearing individuals, I am interested in fair and equitable access. One of the many services CIDI provides is real-time human STTS captioning/transcription. CIDI's mission is to improve the human condition, so when new thoughts, technologies, or applications provide equitable access, we welcome and promote these new discoveries when warranted. As CIDI advises our educational customers, grant funders, and research partners, it was helpful to understand the limitations of ASR during the online surge in course delivery and informational sessions during the COVID-19 pandemic. The results have been incorporated into our plans for best practices.

TARGET AUDIENCE AND RELEVANCE

The target audience for this article has increased during the pandemic to include anyone providing remote

delivery of informational and educational sessions. Public health entities, educators and trainers, researchers, private and public companies, and all providers of remote services will benefit from the insight gained from these best practices. I will provide statistical accuracy ratings for leading ASR systems often used as an accommodation for Deaf and hard of hearing individuals. These results establish best practices for using ASR in its current form and when to use human-provided STTS for access to informational and educational content in online delivery systems.

The relevance of this article is to call attention to our responsibility to consider the impact technological advances may have on users. For example, during the 1990s in the United States, there was interest in moving away from braille for blind individuals and moving to the new screen reading software technology. “Braille has remained vital to the literacy of people who are blind, and it continues to thrive despite the predictions of some to the contrary” (Braille Authority of North America, 2011, p.8). The thought of trading braille materials for screen readers neglects some fundamental issues. When blind individuals read braille, they are more aware of word usage and spelling. A person’s understanding is enhanced by reading independently, and the ability to return to reread a section is readily available. Enhanced memory retention is possible when interacting with braille materials. These benefits demonstrate that new technology (i.e., Screen Reading Software) could not completely replace the former technology. Therefore, there is a continued need for braille documents, and there is a need for screen reading technology.

The same can be said for improvements in transcription. Human-provided Speech-to-Text-Services (STTS) are still important, but with improvements in ASR, ASR produced transcripts used in real-time platforms and recorded media are another way to provide immediate access to spoken speech. Through ASR, people using these technologies can experience automatic, no-cost, or low-cost transcripts. However, considerations must be made in determining when ASR use is appropriate to fulfill access needs for Deaf and hard of hearing individuals.

As a side note, there are some hybrid situations where ASR is used for the initial transcription, but a human STTS provider or collaborative editor listens and corrects in real-time (Akita et al., 2015; Wald, 2018; Wald, 2019). In-progress changes to the real-time transcript are distracting and require some re-training for the Deaf and hard of hearing users to delay reading for a few minutes to allow time for the corrected text to appear. In hybrid real-time situations, where most presentations combine visual presentations with speech, this delay in reading the final version puts the Deaf and hard of hearing individual at a disadvantage for understanding. In this case, a Deaf and hard of hearing individual cannot fully participate in discussions or group interactions.

THE CASE FOR AUDIO ACCESS FOR DEAF AND HARD OF HEARING INDIVIDUALS

The National Deaf Center on Postsecondary Outcomes (2019) affirms that it is appropriate to ask a Deaf or hard of hearing individual what their preferred mode of communication is and honor that specific

request. Possible options may include American Sign Language, real-time Speech-to-Text Services (STTS), such as C-Print software, Typewell software, Communication Access Realtime Translation (CART), or Automatic Speech Recognition (ASR). The goal is to provide successful communication, which allows Deaf and hard of hearing individuals to participate fully.

During the COVID-19 pandemic, some Deaf and hard of hearing users of ASR-produced transcripts stated that they found the use of ASR physically draining because ASR required so much concentration to fill in the blanks or make sense of the transcriptions. Recently a colleague using ASR in an educational setting was frustrated over the missed words and incorrect representation of a speech.

However, some Deaf and hard of hearing users of ASR-produced transcripts have found these transcripts helpful. They noticed errors, but the ASR-produced transcripts met their needs in in low-stakes contexts. A low-stakes setting, determined by the Deaf and hard of hearing individual, may offer a use case for ASR-produced transcripts. An example is a small group planning a family celebration. The vital distinction is that low-stakes situations are settings where consequences are either negligible or easily recoverable. Conversely, high-stakes situations may consist of information leading to an exam, training for work-related tasks, health and safety information, job interviews, *et cetera*. In high-stakes situations, consequences have long-term repercussions.

ASR is promoted as a specialized communication service, which is equivalent to humans (National Deaf Center on Postsecondary Outcomes, 2020). Otter AI (2021) mentions that their service is useful for important conversations and personifies their ASR as a helpful assistant. New opportunities throughout the COVID-19 pandemic provided the ability to learn where ASR-produced transcripts may be successful and situations where they may not be successful. I will discuss my professional experiences with some of the nuances in the outcomes between ASR-produced transcripts and human STTS-produced transcripts.

Unnecessary Utterances

A human STTS provider is trained to provide a true representation of the speaker. Occasionally, human STTS providers may omit unnecessary utterances, such as inarticulates; filler words, such as: uh, um, er, ah, like, right, ok, so, and you know, and other unusual speech patterns. In specific settings, omitting inarticulates aids in understanding by the Deaf and hard of hearing user. The ASR software dutifully represents every sound, every “like,” every “um,” every lead-in utterance, and every side comment. The inarticulates, filler words, or unusual speech patterns may be more distracting than meaningful in a textual representation.

The ASR-produced transcript may match precisely word-for-word in some cases, but meaning may be off. For example, the speaker may have a habit of using a lead-in phrase, as in, “Well, actually,” “Apparently,” “Okay, now,” or other possible phrases. The speaker may establish a pattern of using lead-in phrases that a human STTS provider quickly understands do not add meaning. To alleviate confusion, they will omit them. A human STTS provider example sentence, “Well actually, the earth revolves around the sun,” may become “The earth revolves around the sun.” For a human STTS provider, some loss of

precise word-for-word matching is a professional choice to aid in understanding. For some presenters or speakers, precise word-for-word matching may obfuscate a sentence's meaning.

Best Guess

ASR software is designed to capture every utterance. The ASR transcript will display a word for each utterance, whether it is accurate or not. Oneata et al.'s (2021) study supported that ASR systems use "confidence estimates for a number of downstream tasks: propagating uncertainties for automatic speech translation." Confidence estimates are how ASR software makes a "best guess" for words or phrases. These best guesses are not always noticed by Deaf and hard of hearing users in an informational or educational situation. If the "best guess" is plausible, Deaf and hard of hearing users may absorb the wrong information. Conversely, if the "best guess" is noticeably far from the sentence meaning and Deaf and hard of hearing users are skilled in reading English, they may seek guidance to fill in the missing elements and learn to distrust the information.

Human STTS providers intend to remain faithful to the message without replacing a term misunderstood with a "best guess." Therefore, when a human STTS provider does not hear correctly or recognize a spoken term, they are trained to type a parenthetical, such as "[indistinct]" or "[cannot hear]". In this case, Deaf and hard of hearing users are empowered to seek guidance on the missing elements from the speaker or presenter. Humans are highly capable of understanding unknown speakers in poor audio environments, using arbitrary utterances (O'Shaughnessy, 2008).

O'Shaughnessy (2008) states, "ASR functions well in 'matched conditions,' where the system has been previously trained on all: (1) speakers who would test the system, (2) words that may be used, and (3) possible recording conditions" (p. 2977). The perfect conditions for ASR are rare in most real-time human speech cases.

In many situations, human STTS providers are capable of making sense when mishearing. For example, the phrase "I prefer to have a backup plan," in an ASR-produced transcript may read "I prefer to have a back up land." A human STTS provider may have misheard the word backup, but upon hearing the entire phrase and using context, they can provide the meaningful phrase, "I prefer to have a backup plan." The text represents language, and humans are skilled in managing complex human language nuances.

Human STTS providers can summarize without changing the meaning of a statement. As noted, human STTS providers may omit lead-in phrases, such as "Alright," "So now," "Yeah, well," that may be distracting and do not add to the instructional content. Human STTS providers may also move or omit a side comment that interrupts the focus of the presenter's statement. Human STTS providers can display a more digestible sentence for the user.

Comparative Analysis

We conducted a comparative analysis to determine the accuracy of ASR-produced transcripts in adult educational lecture situations. In addition to Ballenger, the researchers included Matthew Blake, IT Manager, and Kenneth Thompson, Application Developer from the CIDI Information Technology team.

Blake and Thompson searched for leading ASR platforms that process audio files. They selected five options (Microsoft Speech to Text, Google Cloud Speech-to-Text, IBM Watson Speech to Text, AWS Transcribe, and Otter.ai) to obtain a sample of ASR software applications for comparison.

We selected several recorded educational lectures from the CIDI repository representative of adult educational content. The selected lectures were presented in English but varied in topic, presenter, and complexity. We limited the number of recorded lectures to five (5), and we reduced the lecture recording lengths to 30 minutes for a total of 150 minutes of audio recordings containing about 16,000 words. In addition, we created a corrected transcript for each recorded audio file to serve as the baseline transcript. We used the baseline transcript to test the accuracy of the ASR-produced transcripts.

The recorded lecture files were named by their topic and then assigned a random numerical suffix from one through five for that topic. Finally, an independent evaluator submitted the five recorded audio files to the five ASR software services. All transcript collection took place during the first quarter of 2020.

We needed a quantitative comparison to determine the accuracy of the ASR-produced transcripts. First, the differences between each ASR transcript and the corresponding baseline transcript were counted to create a quantitative comparison. We used a custom programmatic tool to automate the comparison, allowing automated word-for-word comparisons. The custom programmatic tool counted the number of words that had to be inserted or deleted in an ASR-produced transcript to be identical to the baseline transcript. An accuracy measure was created by dividing the number of correct words in each ASR-produced transcript by the number of total words in the baseline transcript. The accuracy measure shows a percentage of correctness. A higher number means a more accurate transcript. The accuracy measure of each transcription method is shown in Table 1 in ranked order.

Table 1: Accuracy Measure of Each Transcription Method

Rank Order	ASR Software	Accuracy Measure (%)
1	Otter.ai	88
2	Microsoft Speech-to-Text	86
3	AWS Transcribe	78
4	IBM Watson Speech-to-Text	72
5	Google Cloud Speech-to-Text	69

The ASR-produced transcripts were not fully accurate. Inaccurate transcripts would cause Deaf and hard of hearing users to learn inaccurate information. Therefore, ASR-produced transcripts are not equitable access in high-stakes settings for Deaf and hard of hearing individuals. According to the National Deaf Center on Postsecondary Outcomes (2019), “ASR cannot be relied upon due to high rates of inaccuracy” (The National Deaf Center on Postsecondary Outcomes, 2019).

OUTCOMES AND BENEFITS

An outcome during the COVID-19 pandemic that we learned of was that most in-person informational

and educational sessions shifted to remote with little or no thought to providing appropriate accommodations for Deaf and hard of hearing individuals. Likewise, the rushed conversion to ASR-produced transcripts provided during the COVID-19 pandemic was not a quality accommodation for informational and educational settings.

The benefit is that there are some appropriate uses for ASR-produced transcripts. The use is in low-stakes settings. ASR-produced transcripts provide a path to no-cost or low-cost options for this particular use. It is appropriate to ask the Deaf and hard of hearing participants their preference for human STTS-provided captions/transcription or ASR-produced transcription. Understanding that there is a distinct difference in access for high-stakes settings ensures that the Deaf and hard of hearing individual has equitable access for informational and educational communication.

Additionally, this article provides a framework for future studies of the efficacy of ASR software used as an accommodation.

CONCLUSION

At CIDI, we continue to provide human-provided STTS access to remote informational or educational sessions, as ASR software does not provide equitable access for Deaf and hard of hearing individuals. This decision was applied to our involvement with the Centers for Disease Control and Prevention, resulting in accessible human-provided STTS real-time captioning/transcription and American Sign Language for public-facing informational and educational sessions.

For Deaf and hard of hearing participants, ASR-produced transcripts may be helpful for low-stakes information sessions, especially when speaker/presenters have multiple means of representing what they are sharing. The decision of whether to supply human-provided STTS access or ASR-provided access should be determined by the Deaf and hard of hearing user's preference.

As discussed, unnecessary utterances are distracting in a textual representation, and incorrect best guesses change the words and alter the meaning. Perhaps ASR developers could provide a specific ASR application for Deaf and hard of hearing users with the option to omit unnecessary utterances and mark text below a certain confidence level.

Similar to braille for blind individuals, there will remain situations where human STTS-produced transcripts are essential for Deaf and hard of hearing individuals. In audio environments that are less than ideal, human's contextual abilities are much more astute than ASR. For example, some speakers with different speech patterns or accents are better understood by a human STTS provider. High-stakes informational or educational settings where Deaf and hard of hearing individuals will be responsible for understanding and applying the information require human STTS providers. Human STTS providers should be supplied when the reliability of the information is essential.

Determining whether or not in-person or remote informational or educational sessions are accessible, the U.S. Department of Justice (2014) states, “The goal is to ensure that communication with people with these disabilities is equally effective as communication with people without disabilities (p.1). Future studies are necessary to assess how new transcription techniques could benefit Deaf and hard of hearing users and how providing access with ASR could be helpful in certain situations. Remember to seek preferences of audio access from the Deaf and hard of hearing individuals included in your remote environments (The National Deaf Center on Postsecondary Outcomes, 2019).

DECLARATIONS

The findings and conclusions in this report are those of the author(s) and do not necessarily represent the official position of the Centers for Disease Control and Prevention or ATIA. Development of these materials was supported in part by a grant from the CDC Foundation, using funding provided by its donors. The materials were created by the Center for Inclusive Design & Innovation (CIDI), Georgia Tech. The CDC Foundation and Centers for Disease Control and Prevention (CDC) provided subject matter expertise and approved the content. The use of the names of private entities, products, or enterprises is for identification purposes only and does not imply CDC Foundation or CDC endorsement.

REFERENCES

Akita, Y., Kuwahara, N., & Kawahara, T. (2015). Automatic classification of usability of ASR result for real-time captioning of lectures. *Proceedings of APSIPA Annual Summit and Conference* (pp. 19-22). APSIPA. http://sap.ist.i.kyoto-u.ac.jp/EN/bib/intl/AKI-APSIPA15.pdf

Alberti, C., & Bacchiani, M. (2009, December 4). Automatic captioning in YouTube. *Google AI Blog* [Web log post]. https://ai.googleblog.com/2009/12/automatic-captioning-in-youtube.html#:~:text=On%20November%2019%2C%20we%20launched,captions%20for%20videos%20on%20YouTube

Alsalamah, A. (2020). Using captioning services with deaf or hard of hearing students in higher education: a systematic review. *American Annals of the Deaf, 165*(1), 114–127. https://doi.org/10.1353/aad.2020.0012

Barnett, S., McKee, M., Smith, S. R., & Pearson, T. A. (2011). Deaf sign language users, health inequities, and public health: opportunity for social justice. *Preventing Chronic Disease, 8*(2), A45. Retrieved from https://www.cdc.gov/pcd/issues/2011/mar/pdf/10_0065.pdf

Bogdan, I. (2019). Evaluating Google speech-to-text API’s performance for Romanian e-learning resources. *Informatica Economică, 23*(1), 17–25. https://doi.org/10.12948/issn14531305/23.1.2019.02

Braille Authority of North America. (2011). The evolution of braille: can the past help plan the future? (Part 1). http://www.brailleauthority.org/article/evolution_of_braille-part1.pdf

Buchanan, B. (2005). A (very) brief history of artificial intelligence. *AI Magazine, 26*(4), 53. https://doi.org/10.1609/aimag.v26i4.1848

Ekbia, H., & Nardi, B. (2014). Heteromation and its (dis)contents: The invisible division of labor between humans and machines. *First Monday, 19*(6). https://doi.org/10.5210/fm.v19i6.5331

Ellcessor, E. (2020). Three vignettes in pursuit of accessible pandemic teaching. *Communication, Culture and Critique, 14*(2), 324–327. https://doi.org/10.1093/ccc/tcab010

Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020, March 27). The difference between emergency remote teaching and online learning. *Educause Review.* https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning

Hoglind, T. (2018). Healthcare language barriers affect deaf people too. *Boston University School of Public Health.* https://www.bu.edu/sph/news/articles/2018/healthcare-language-barriers-affect-deaf-people-too/

National Association of the Deaf (2021). Position statement on health care access for deaf patients. https://www.nad.org/about-us/position-statements/position-statement-on-health-care-access-for-deaf-patients/

National Deaf Center on Postsecondary Outcomes (2019). *Effective communication tip sheet.* https://www.nationaldeafcenter.org/resource/effective-communication

National Deaf Center on Postsecondary Outcomes (2020, October 27). *Auto captions and Deaf students: Why automatic speech recognition technology is not the answer (yet).* Retrieved April 23, 2021 from https://www.nationaldeafcenter.org/news/auto-captions-and-deaf-students-why-automatic-speech-recognition-technology-not-answer-yet

Oneata, D., Caranica, A., Stan, A., & Cucu, H. (2021). An evaluation of word-level confidence estimation for end-to-end automatic speech recognition. *IEEE Spoken Language Technology Workshop (SLT),* (pp. 258–265). IEEE. https://doi.org/10.1109/SLT48900.2021.9383570

O'Shaughnessy, D. (2008). Automatic speech recognition: history, methods and challenges [Invited paper]. *Pattern Recognition, 41*(10), 2965–2979. https://doi.org/10.1016/j.patcog.2008.05.008

Otter AI. (2021). Otter is where conversations live. Retrieved May 24, 2021 from https://otter.ai/

Parton, B. (2016.). Video captions for online courses: Do YouTube's auto-generated captions meet deaf students' needs? *Journal of Open, Flexible and Distance Learning, 20*(1) 8–18. https://files.eric.ed.gov/fulltext/EJ1112346.pdf

Section508.gov. (n.d.). https://www.section508.gov/create/accessible-meetings/

U.S. Access Board. *Revised 508 standards and 255 guidelines.* Retrieved April 5, 2022 from https://www.access-board.gov/ict

U.S. Department of Justice. (2014). *Effective communication.* www.ada.gov/effective-comm.htm

Wald, M. (2018, June 23 - 25). Using speech recognition transcription to enhance learning from lecture recordings. *International Conference on Education and New Developments* (pp. 111–115). University of Southampton. https://eprints.soton.ac.uk/419608/

Wald, M. (2019, March 17). Enhancing learning through accessible lecture recordings. *Media and Learning.* https://media-and-learning.eu/type/featured-articles/enhancing-learning-through-accessible-lecture-recordings/

World Health Organization. (2020). *Deafness and hearing loss.* https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss