Accuracy of Speech-to-Text Captioning for Students Who are Deaf or Hard of Hearing

Pam Millett, PhD
Associate Professor
Deaf and Hard of Hearing Program
Faculty of Education
York University
113 Winters College
4700 Keele St., Toronto, Ontario, Canada
M3J 1P3
pmillett@edu.yorku.ca
416 578-6001

ABSTRACT

Speech-to-text technology (also referred to as automatic speech recognition, or ASR) is now available in apps and software, offering opportunities for deaf/hard of hearing students to have real time captioning at their fingertips. However, speech-to-text technology must be proven to be accurate before it should be considered as an accommodation for students. This study assessed the accuracy of eight apps, software and platforms to provide captions for i) a university lecture given by a native English speaker in real time ii) a video of the lecture, and iii) a conversation between 3 students in real time, using real speech under controlled acoustical conditions. Accuracy of transcribed speech was measured in two ways: a Total Accuracy score indicating % of words transcribed accurately, and as a Meaning Accuracy score, which considered transcription errors which impacted the meaning of the message. Technologies evaluated included Interact Streamer, Ava, Otter, Google Slides, Microsoft Stream, Microsoft Translator, Camtasia Studio and YouTube. For the lecture condition, 4 of 5 technologies evaluated exceeded 90% accuracy, with Google Slides and Otter achieving 98 and 99%% accuracy. Overall accuracy for video captioning was highest, with 5 of 6 technologies achieving greater than 90% accuracy, and accuracy rates for YouTube, Microsoft Stream and Otter of 98-99%. Accuracy for captioning a real time conversation between 3 students was greater than 90% for both technologies evaluated, Ava and Microsoft Translator. Results suggest that, given excellent audio quality, speech-to-text technology accuracy is sufficient to consider use by postsecondary students.

Key Words: **Speech-to-text, captioning, deaf or hard of hearing**

INTRODUCTION

Since the early 1980s, captioning has been recognized as an important technology for accessibility for individuals who are deaf or hard of hearing (Block & Okrand, 1983). In recent years, the use of automated speech-to-text technology, in which computer software uses voice recognition to transcribe speech into print, offers the potential to provide "on the go" captioning across a wide variety of communication, occupational, and educational situations. Businesses are interested in speech-to-text captioning to provide access for employees, clients or customers who are deaf or hard of hearing (Ajami, 2016; Oberteuffer, 1995; Qiu & Benbasat, 2005; Vajpai & Bora, 2016; Venkatagiri, 2002). Consumers are looking for inexpensive, easy to use automatic captioning for everyday interactions (for example, around a dinner table or at church) (Morris & Mueller, 2014). One of the most exciting applications of this technology, however, is for classroom learning. Administrators at colleges and universities, and providers of educational programming or professional development, are interested in speech-to-text captioning to provide captioning in classroom lecture situations and for online learning materials (Wald & Bain, 2008). There is a large body of research indicating that captioning improves comprehension for deaf students, students with learning disabilities and English Language Learners (see Gernsbacher, 2015 and Perez, Van Den Noortgate & Desmet, 2013 for reviews), yet the provision of captioning in education continues to be primarily on a case-by-case basis (Kent, Ellis, Latter & Peaty, 2018).

Accessibility Challenges in Today's Postsecondary Education Learning Environment

Recent research suggests that 75 to almost 90% of students with hearing loss are being educated in regular classrooms with hearing peers, rather than in congregated classrooms or schools for the deaf, using spoken language rather than sign language (CRIDE, 2018; Reed, Antia, & Kreimeyer, 2008). Many of these students will go on to postsecondary education, and should expect to have appropriate accommodations in place at any institution they choose to attend. However, postsecondary education represents a different learning environment, and students may require more or different accommodations than were needed or provided in elementary or secondary school (Powell, Hyde, & Punch, 2013). Cawthon et al. (2013) reported that the most common accommodations reported to be used in secondary and postsecondary education by students with hearing loss were sign language interpreters, followed by notetakers, sound amplification, captioned media (e.g. videos) and extended time for tests. Real time captioning was one of the least used accommodations, yet it is a service which provides accessibility for students who have a variety of communication preferences (including students graduating from mainstream education who do not use sign language).

Challenges for In-class Lectures and Learning Activities

Large class sizes and poor room acoustics are often seen in postsecondary classrooms, creating suboptimal listening environments for hearing students due to the detrimental effects of distance, noise, and reverberation (Aish, Jain, & Jain, 2019; Ibrahim, 2020; Crandell, Kreisman, Smaldino, & Kreisman, 2004; Madbouly, Noaman, Ragab, Khedra, & Fayoumi, 2016). Potential barriers to learning are not confined to the physical classroom listening environment, however. The typical college or university student is likely to take five or more classes per semester, encountering five different instructors with unique teaching and speaking styles, five different acoustic environments, and five different cohorts of classmates of varying size. The course may consist of many hours of an instructor lecturing to a large classroom or it might include a significant number of requirements for student collaboration in small group work. Instructors routinely take advantage of multimedia resources (such as YouTube videos) which can be difficult for a student with hearing loss to access. The inclusion of projects and activities that encourage student collaboration is an important pedagogical practice, but small group work can be difficult for a student with hearing loss if noise levels are high. Additionally, group work might also require collaborating outside of class time, with classmates who are in a variety of geographical locations, requiring the use of videoconferencing, Facetime, videochat, or other telecommunication means that may not provide the necessary auditory and visual cues needed for comprehension. Finally, in the year of the COVID-19 pandemic, on some campuses, online learning is the only option provided, creating a host of potential barriers for students with hearing loss (Millett, 2019; Millett, 2010; Millett & Mayer, 2010).

To minimize some of these challenges, an effective classroom accommodation can be the provision of real-time captioning using a Computer Assisted Real-time Translation (CART) captionist, where a captionist is present during class (either in person or remotely) and produces an exact transcript of everything heard in the classroom in real time. Remote captioning, where the captionist is located offsite but listening in through a speakerphone, Skype, etc., reduces some of the cost. However, providing CART during a live lecture has technical challenges which are difficult to address in a typical postsecondary classroom. CART requires the use of a very good microphone in the classroom, so that the captionist is able to hear clearly enough to caption accurately. Poor audio quality, signal dropping, background noise, or speakers moving away from the microphone will compromise the captionist's ability to hear and transcribe accurately.

Administrative resources to contract and schedule captionists for all classes on the student's timetable are also required, to ensure that CART is always available when needed. If remote captioning is used, the institution's instructional technology department needs to be involved create a setup that provides sufficient audio quality to the offsite captionist. Instructors need to be trained on how to use the audio setup and on principles of best practice (for example, the timing of breaks). Because CART requires specialized hardware, software, and skills, it is an expensive accommodation in itself, as well as requiring significant administrative resources to ensure the service is available for each student, for each class. Expense and complicated administrative logistics have meant that this extremely effective accommodation has not been routinely available for most deaf and hard of hearing students (Kawas, Karalis, Wen, & Ladner, 2016).

Online Learning and Web-based Learning Materials

Challenges with online learning for students with disabilities have been documented in the literature (Burgstahler, 2015; Fichten, Asuncion, Barile, Fossey, & Simone, 2000), although Long, Vignare, Rappold, and Mallory (2007) found that students who are deaf or hard of hearing reported positive aspects of online learning as well. Online learning opportunities in postsecondary education typically include extensive use of web-based audio and video materials to improve learning and student engagement, which creates challenges for students with hearing loss. CART can be a feasible accommodation in an online course, but can be expensive and logistically complicated to arrange, given the diversity of students in a postsecondary institution enrolled in a wide variety of programs and courses. While providing a written transcript of an audio or video file does provide some accessibility, it is impossible for a student to watch a video while simultaneously reading a written transcript. Jelinek Lewis & Jackson (2001) in fact, found that deaf students' comprehension of captioned material was best when captions were provided in the video, and poorest based on reading the transcript alone.

Increasingly, lecture capture, where an instructor videotapes a lecture or tutorial for asynchronous viewing by students, is being used at the postsecondary level for online learning applications

(O'Callaghan, Neumann, Jones, & Creed, 2017). With extensive moves to online learning this year created by the COVID-19 pandemic, it is likely that this method of content creation is increasingly common. This poses potential problems for students with hearing loss since a recorded lecture typically does not provide the same experience as sitting in class: the audio may not be as clear, and the quality of the video recording of the instructor may not allow for lipreading. Lecture capture files can be captioned in postproduction fairly easily if there is infrastructure in place for captionists to quickly access the files; there are a variety of commercial companies that provide this service. This solution requires a fairly significant initial investment in time and money by the IT department. However, once implemented, video captions can be provided quite seamlessly and, because these services typically use speech-to-text technology supplemented by subsequent human editing, accuracy is very high. Providing captions for lecture capture files is very feasible, but requires an institutional commitment to allocate the necessary staff, money and resources.

A more significant captioning problem is presented when instructors wish to use web-based video. While a standard accommodation recommendation has always been "use captioned videos when possible," the reality is that instructors use less video material of this type (for example, showing something from a DVD in class) and more video material that they find on websites or through applications such as YouTube or Vimeo. Some applications, such as YouTube, offer a captioning service that uses voice recognition software to transcribe the audio in real time. Although there has been some research on YouTube for specific captioning challenges such as dialect (Tatman, 2017), data on the reliability and accuracy of these applications (other than what is reported in marketing materials) is extremely scarce. Captioning quality of web-based videos can range from excellent to poor to completely unusable. Furthermore, copyright restrictions make it difficult to provide captioning for video material that is owned by another user.

While there is video editing software that allows an instructor to caption a video (such as Camtasia Studio), it is unlikely that the average instructor will have the financial or technological resources to do this. In fact, Hinds and Mather (2007) found that instructor attitudes towards professorial accommodations (i.e., situations which required the professor to change or do something extra or different) were not favorable. Access to captioned audio and video material used in class therefore remains a significant problem for students who are deaf or hard of hearing. In the U.S., the National Association for the Deaf has recently won lawsuits against institutions, notably Ohio State University and University of Maryland, which did not provide captions for publicly displayed materials (such as football scoreboards) (Grasgreen, 2013).

Speech-to-text technology may offer cost effective and user friendly technologies for students and instructors to have the ability to access real time captioning at their fingertips through an app or software in any situation where they are experiencing communication difficulties for any of the activities described above – assuming that the captions are accurate. How, though, should interested users of speech-to-text captioning find and evaluate this technology for their own use? Wald and Bain (2008) note that, in order to be effective, speech to text technology needs to be:

> be completely transparent to the speaker (i.e., it would not require training and no microphone would need to be worn by the speaker), be completely transparent to the "listener" (i.e., it would not require the user to carry any special equipment around), recognise the speech of any speaker (even if they had a cold or an unusual accent), recognise any word in any context (including whether it is a command), recognise and convey attitudes, interest, emotion and tone, recognise the speaker and be able to indicate who and where they are, [and] cope with any type or level of background noise and any speech quality or level. (p. 438)

This is a high bar for technology to reach. There are increasing numbers of apps available to either purchase or use for free, but there is little research on accuracy and effectiveness available for consumers.

Research on the Use of Speech-to-Text Technology for Captioning in Classrooms

The body of literature on the accuracy of speech-to-text technology has largely used static and controlled speech conditions, with little research on the accuracy, feasibility, and efficacy of speech-to-text technology in either physical or virtual classrooms. In fact, there are no guidelines on acceptable accuracy rates even from regulatory bodies or legislation such as the FCC and the 21st Century Video Accessibility Act (CVAA) (3Play Media 2020). The majority of the research in classrooms comes from the Liberated Learning Consortium, a partnership between Canadian and international universities and IBM research (Bain, Basson, & Wald, 2002; Ranchal et al., 2013). The Consortium's research indicated that in a real classroom, speech-to-text software was able to provide accuracy rates of 85% or better for only 40% of instructors. High accuracy rates were only seen when instructors were required to first train the software to their voices, and then use wireless microphones to improve audio quality. Although the researchers reported that students generally liked the captioning provided (as long as accuracy was >85%), the students surveyed had typical hearing and therefore were receiving the intact message auditorily in addition to looking at the captions, which likely minimized the impact of captioning errors on comprehension.

Transcribing multiple voices in a classroom discussion is a situation which is difficult even for human captionists. While Wald (2008) reported some success in using Liberated Learning software in a meeting type setting with multiple speakers, each speaker needed to have their own microphone, a solution which is not feasible in a postsecondary classroom with several hundred students.

Suggestions for accuracy improvement have included having an editor correct the captioning or having someone re-voice the lecturer's message more clearly, neither of which would be feasible in a real classroom setting. Some researchers have suggested that using crowdsourcing or volunteers could be a potential avenue for having a human correct the speech-to-text transcript (Takagi, Itoh & Shinkawa, 2015; Wald, 2006, 2011). However, the feasibility of implementing this strategy in a typical postsecondary classroom seems low, and the limited body of research on speech-to-text technology in real classrooms continues to highlight difficulties with accuracy that are not easy to solve.

The purpose of this study was to determine the accuracy of a variety of commercially available speech-to-text apps and software to provide captions for a variety of learning activities. Several situations that might arise in a typical postsecondary course were considered: i) listening to a live lecture, ii) accessing online video and audio materials as part of the coursework and iii) interacting with other students in a small group work setting. Informal qualitative analysis was also conducted to identify potential issues from the perspective of the user.

METHOD

The focus of the study was to evaluate speech-to-text technology accuracy in a classroom in real time, with an instructor delivering a lecture in the rate, pace, and style that he/she would typically use, while keeping a controlled acoustical environment. Using an actual real-time lecture would have been the most realistic and representative, but the noise levels, distance, and reverberation in most postsecondary classrooms likely would have resulted in significant deterioration of the captioning accuracy. While evaluation in a standard classroom is important to do, the goal of this exploratory study was to evaluate captioning accuracy in a live lecture situation within a controlled acoustic environment to minimize the effects of noise, distance, and reverberation.

Materials for the live lecture condition

A 15 minute excerpt from a previously recorded lecture in a university course in Educational Audiology was used as the script for live voice input to speech-to-text captioning technologies. An exact transcript was created by reviewing the recording, including filler words and false starts (see Appendix A for transcript). The researcher then used the correct transcript to give the same lecture again, but this time in a quiet, acoustically treated classroom with no noise and using a microphone, with captioning apps and software running on a variety of devices. The intent was to create a naturalistic presentation of the lecture using live voice as input to the various speech-to-text technologies (rather than a recording), while incorporating some standardization so that new technologies could be evaluated in future using the same stimuli. The classroom used was designed specifically for deaf and hard of hearing students, and is fully carpeted, with acoustic panels up to approximately 70% of the wall weight, and acoustic panels on the ceiling. Unoccupied noise levels for the classroom were measured at 30 dBA.

Most speech-to-text applications recommend using a Bluetooth headset paired to the device on which the speech-to-text software is running, for best sound quality. Often, students with hearing loss are already using a specialized version of this setup, with the instructor using a specialized microphone which transmits his/her voice directly to the student's hearing aids or cochlear implant. In a real classroom, the setup for using speech-to-text technology would involve connecting the student's existing hearing device to the tablet, laptop or cellphone on which the app or software is located. The audio configuration for this study used Phonak Roger Touchscreen transmitter worn by the researcher with a Phonak Roger MyLink receiver coupled to the device running the speech-to-text technology (laptop, tablet, or cellphone) using a TRS/TRRS cable + 3.5/2.5 mm adapter. This simulates what would happen in a real classroom, where the instructor would be wearing a Phonak transmitter, and the student would have a Phonak receiver connected to his/her phone or tablet. The student would then hear the instructor through his/her hearing aids or cochlear implant, while the phone or tablet would simultaneously receive the speech signal for captioning. In this study, multiple Phonak receivers connected to a variety of devices were used, to ensure that all apps/software received the same speech signal at the same time, under the same conditions. The transcript produced by each app/software was then saved, and an accuracy rate was calculated by comparing the speech-to-text transcript to the correct transcript.

Both a CART transcript and professional post production captioning were also created at the time that the original lecture was given. The accuracy rates of the CART and the post production transcripts were also calculated by comparing them to the correct transcript.

Materials for video captioning

The lecture was simultaneously recorded on Audacity, converted to an MP4 file and then used as the input for video captioning. This was intended to simulate a lecture capture application, or a situation where an instructor might record a lecture or tutorial him/herself, for students to view outside of class. The transcript produced by each app/software was then saved, and an accuracy rate was calculated by comparing the app transcript to the correct transcript.

Materials for small group work captioning

Clearly, the language, speaking style and rate of an instructor giving a lecture is different from the language and conversation style of a conversation between students working in a small group. Therefore, a script was written for a group conversation with three students regarding a biology course project on the discussion topic "Name three species (common and scientific name) that have traits that surprised you. What is the advantage of this trait, or do you think it provides a disadvantage? How do you think this trait could have developed from its ancestors?" (see Appendix B for full transcript). The researcher and two university students downloaded captioning apps to their individual iPhones, set up a group conversation per the app instructions, and used the script as input to the speech-to-text captioning technologies. The transcript of the small group conversation for each app was saved, and then accuracy rates for each were calculated.

Apps, Software, and Platforms Evaluated

A wide array of captioning, dictation, and speech-to-text apps can now be found on the internet, on the App Store and on Google Play. A list of apps and software to be evaluated for this study was first generated from Google, iTunes, and Google Play using the search words "captioning," "speech to text," "automatic speech recognition", "voice to text," "automatic captioning," and "speech-to-text." Each app/software was then evaluated based on a set of inclusion and exclusion criteria. Apps or software specifically marketed for classroom captioning or use by individuals who are deaf or hard of hearing, were included.

Apps, software, and platforms excluded included those intended, or determined to be, for dictation only (e.g., SpeechNotes, Google Keyboard, Transcribe). Other technologies were excluded because they would be difficult or expensive for a typical instructor to implement. For example, Google Cloud Speech-to-Text is a cloud-based live voice and audio file transcription (https://cloud.google.com/speech-to-text/). The live voice version was considered to not be a feasible option for educational use as the time limit for real time captioning is only one minute, and the audio file transcription was felt to not be cost effective or user friendly enough to be recommended for school use. Kapwing, an online video editing service, was also considered but excluded. It provides automatic captioning for an uploaded video, but adds the company logo and deletes content after 7 days unless a monthly subscription is purchased (https://www.kapwing.com/). Finally, apps that had restricted or limited use by teachers or students were excluded (for example, Clips for iPhone provides speech-to-text captioning called LiveText, but only for videos recorded on the app on an iPhone) .

Live lecture condition.

Table 1 provides a summary of the captioning apps and software that were used in this study to evaluate speech-to-text technology in the live lecture condition. Many of the technologies can be used on different devices (e.g., iPad, iPhone, Android phone, Android tablet, PC, or laptop). Only Ava and Interact Streamer are described specifically as being recommended for deaf or hard of hearing for classroom use, but other technologies included are capable of being used for real time captioning of a lecture, even if not specifically marketed to individuals with hearing loss. For example, the option to tweak Google Slides to provide automatic captioning in real time is not widely described in Google Slides informational material; however, it was included in this study based on reports from American and Canadian educational audiologists and teachers of the deaf and hard of hearing that they were trying out this application with students in classrooms.

While there is no evidence to suggest that there would be differences in captioning accuracy depending on the device on which the app, platform, or software was used, this was confirmed by evaluating some technologies on multiple devices. A comparison of captioning accuracy for the same speech-to-text technology used on different devices did not indicate any differences in captioning accuracy. For this condition, the CART transcript was considered the "gold standard."

Table 1. Technologies evaluated for the live lecture condition

App/software	Version	Description and Notes	Devices Used
CART	N/A	CART transcript was done during live lecture	N/A
Ava https://www.ava.me	2.0.9	App available for Android and iOS. Described as offering "on the go" captioning in a variety of settings, including work, school, and social situations. Note: Accuracy level set to default (Auto 95%); Curse Words Filter set to No curse words.	iPhone S iPad Pro MacBook Air laptop
Interact Streamer website www.streamer.center	N/A	Described as a "captioning, translation, messaging and document sharing website. Conversations with friends, conference room meetings, classroom discussions, webinars and religious services are all examples of how Streamer™ is being used every day throughout the world." Users can access captioning through the website or through the app.	MacBook Air laptop
Interact Streamer app	2.2.0	Same technology as on the website. Available for Android and iOS.	iPhone iPad Pro
Google Slides (Presenter Mode) https://www.google.com/slides/about/	N/A	Available for free as part of the office suite in Google Drive. Software will provide real time captioning by opening a blank powerpoint slide, clicking "Present," and then Ctrl + Shift + c. Only works with Chrome browser, does not work in the Google Slides app.	MacBook Air laptop (software) Surface Pro tablet (software)
Otter Otterai.com	2.1.30.584		iPad Pro MacBook Air laptop (software)

Video captioning condition.

Table 2 provides a summary of the apps/software used to evaluate the accuracy of video captioning. The "gold standard" was considered to be that provided by a commercial captioning company, which was provided for the original lecture. In this study, the company used was 3Play, a California-based company that uses speech-to-text captioning plus human editing (https://www.3playmedia.com/).

Table 2. Technologies evaluated for the video captioning condition

App/software	Version	Description and Notes
3Play	N/A	
YouTube (www.youtube.com)	N/A	YouTube offers an option to click on CC, and choose "English (autogenerated)" for captions
Microsoft Stream (https://products.office.com/en-ca/microsoft-stream)	N/A	Stream is a platform included in the Microsoft Office 365 suite, and is intended to allow users to create and share video for meetings and online learning.
Camtasia Studio (https://www.techsmith.com/)	Camtasia Studio 8	Camtasia Studio is video content creation and editing software developed by TechSmith which includes a Speech to Text option to create video captions.
Ava https://www.ava.me	2.0.9	App available for Android and iOS. Note: Accuracy level set to default (Auto 95%); Curse Words Filter set to No curse words.
Google Slides https://www.google.com/slides/about/	N/A	Available for free as part of the office suite in Google Drive. Software will provide real time captioning by opening a blank powerpoint slide, clicking "Present," and then Ctrl + Shift + c. Only works with Chrome browser, and only does not work in the Google Slides app.
Otter Otterai.com	2.1.20.584	

Small group conversation condition.

Two apps were identified that seemed to have the potential to be used for small group work, as they allow multiple users to easily download the app onto their individual devices, and then use their own phone or tablet as a microphone: Ava and Microsoft Translator (https://translator.microsoft.com/). Microsoft Translator is a web-based app developed for language translation but can also provide captioning when the translation is set to "English-to-English." Once the user has logged in, a "chat room" interface is seen, which can be used with one or more users in a conversation. Microsoft Translator requires the speaker to press and hold an icon while talking, which would be awkward and tiring to do in a lecture format. However, it would work well in a small group situation where speakers simply touch the icon when they want to talk, and release it when they are finished talking. Ava also allows multiple users to login to a chat room interface, and the app then provides real time captioning of each user's voice.

Scoring

Captioning accuracy is typically reported as either a Word Error Rate (WER), which refers to the total number of errors in the transcript, or as an Accuracy percentage, referring to the total percentage of speech transcribed accurately. Errors can also be coded qualitatively in a variety of ways (e.g., omissions, additions, substitutions, phonetic errors, etc.). VanZant (2015), for example, looked at real time captioning errors in broadcast news programs, and identified 15 different types of captioning errors. She concluded that "Knowing which types of errors are the most common can lead to a better understanding of how to prepare deaf students to face these challenges" (p. 19). However, focusing only on identifying type of error does not take into account that the impact of an error on user comprehension depends on the linguistic context. VanZant found that spelling errors were most common; however, that does not mean that spelling errors necessarily impact comprehension. An app might have a relatively high number of errors, but if the errors do not change the meaning of the message and do not affect comprehension, a relatively poor accuracy rate might not actually represent a problem for users. On the other hand, an app might have a relatively small number of errors, but if they are errors that dramatically degrade meaning and comprehension, a high accuracy rate is still a problem.

In this study, all errors were counted but not coded as omissions, additions, or substitutions. Instead, further qualitative assessment was conducted with respect to impact on meaning. The best and most accurate determination if a particular error impacts meaning, and therefore comprehension, is by a human listener who is a competent user of the language. As such, a second score was calculated, Meaning Accuracy, in which errors that did not impact meaning were removed. These errors often represented morphological markers, function words, plural, articles, or extraneous words, but not always. For example, "it becomes a problem" incorrectly transcribed as "this becomes a problem" was counted as an error for the Total Accuracy score (since it is an actual pronoun error). However, from the perspective of a user, it was not considered to impact meaning in this context, and was not counted as a Meaning error. Homophonous errors were counted as Meaning errors, however.

Because Meaning Accuracy is a subjective measure that has not been used previously in the literature, inter-rater reliability was also calculated. A university undergraduate student (representative of a user of real time captioning) was provided with three randomly chosen transcripts that showed both the errors and the correct language. The student was asked to identify the errors that did not affect meaning. The Meaning Accuracy scores identified by the researcher and the student were compared. Inter-rater reliability using Cohen's Kappa was found to be 94%, suggesting that there is good agreement amongst competent users of the English language as to what kinds of errors compromise reader comprehension.

RESULTS

Results are provided for each condition (lecture, video, and small group conversation), with Total Accuracy and Meaning Accuracy scores presented separately, along with qualitative observations on punctuation. Table 3 provides the results for each technology for the lecture condition.

Live lecture condition

Table 3. Real time captioning accuracy for the lecture condition

App or Software	% Total Accuracy	% Meaning Accuracy	Observations on Punctuation	Other Notes
CART	99.8	99.8	Fully correct	
Ava	87.2	87.9	Periods only	Captions presented as a transcript; transcript can be saved.
Interact Streamer website	93.6	94	Misleading; uses - ! : ; ? . often in wrong place	Captions presented as a transcript; transcript can be saved
Interact Streamer app	93.6	94	Misleading; uses - ! : ; ? . often in wrong place	Captions presented as a transcript; transcript can be saved.
Google Slides	97.7	97.7	None	Captions presented under PowerPoint slides.
Otter	99.7	99.7	Generally correct	Captions presented as a transcript; transcript can be saved; identifies different speakers

The most accurate transcripts were provided by Otter. The second most accurate transcript was produced by Google Slides, despite the fact that automatic captioning is not a commonly used, or even frequently described, feature of Google Slides. Interact Streamer, which was developed for real time captioning in a classroom for deaf or hard of hearing students, produced accuracy rates of greater than 90%. However, although it included much greater use of punctuation, the punctuation was frequently incorrect and misleading. The example below demonstrates an Interact Streamer transcription with both semantic and punctuation errors, and then with punctuation errors only. As the example shows, readability clearly continues to be impacted even when all of the actual words are correct:

She said: well, here's a dated audiogram and the Ada. Ian Graham had thresholds at about 60 DB right. If you hadn't ate it on your Gram at 60 DB at four thousand Hertz, do you think they could hear an ass? No here's! The 8th at audiogram I just watched her do a 6lb test.

If the word errors are fixed, but the punctuation errors are maintained, the script reads as:

She said: well, here's his aided audiogram and the aided. audiogram had thresholds at about 60 DB right. If you had an aided audiogram at 60 dB at four thousand Hertz, do you think they could hear an s? No here's! the aided audiogram I just watched her do a 6 sound test.

While "type" of error (e.g., grammatical category or phonetic similarity) was not specifically coded in this analysis (as the focus of the data analysis was on meaning change), it is worth noting that one type of error did particularly stand out. Homophonous errors (for example, "what can a child here?" instead of "what can a child hear?") occurred quite frequently in all speech-to-text technologies. From the perspective of a user of captioning, homophone errors are essentially auditory errors, requiring the reader to repeat the word internally or out loud, in order to identify the correct word, something which may be difficult for many students who are deaf or hard of hearing.

Other errors demonstrated speech-to-text's relatively poor prediction algorithms for lecture applications. Examples of errors include: "Some idiot 2 feet away" for "So maybe at 2 feet away;" "Shepherd's Chapel" for "for example;" "dude" for "student;" "six hour text," "six pound test," or "sick sound test" for "six sound test;" and "thinking varmint," "listing apartment," and "listing of varmint" for "listening environment." It would arguably be impossible for a user to use context to repair the captioning error and understand the meaning of the sentence with errors such as these. At times, the text produced was so far from the original to be unrecognizable, such as in "the listening environment gets exponentially more difficult as soon as babies get mauled up" for "the listening environment gets exponentially more difficult as soon as babies get mobile."

Video captioning condition

Table 4 provides accuracy results for speech-to-text technologies for video captioning, along with results provided by the post-production company (3Play).

Table 4. Real time captioning accuracy for the video condition

Platform	% total accuracy	% A (MM)	Punctuation	Notes
Post production captioning company (3 Play)	99.9	99	Fully correct	Time coded, captions presented under video.
YouTube	98	98.4	Contractions only	Time coded, captions under video.
Microsoft Stream	98.7	98.7	None	Time coded, captions on right hand side of video.
Ava	92.7	93	Occasional periods	Transcript, not time coded.
Google Slides	95.4	96.1	Contractions only	Transcript presented as speaker's notes below slides.
Interact Streamer	89.8	92.8	Misleading	Transcript.
Otter	99.7	99.7	Generally correct	Transcript; can be exported as an srt or vtt file

Otter provided the most accurate transcript, followed by Microsoft Stream and YouTube which all produced accuracy rates which were close to the gold standard, post production commercial captioning. Ava, and Google Slides produced slightly lower accuracy in this uploaded video context, than in transcribing live speech.

Small group conversation condition

Table 5 provides the results for speech-to-text technologies for the small group conversation condition. Both apps were more than 90% accurate, although Microsoft Translator provided slightly better results.

Table 5. Real time captioning accuracy for the small group conversation condition

Platform	% Total Accuracy	% Meaning Accuracy	Punctuation	Notes
Microsoft Translator	94.1	96.9	None	Time coded, captions presented under video.
Ava	92.7	93.4	Occasional periods used	Transcript, not time coded.

Because of the small sample size for each condition, statistical comparisons of accuracy rates were not appropriate. Nevertheless, it is worth noting that there were very few differences between Total Accuracy, and Meaning Accuracy, scores. Interact Streamer and Microsoft Translator were the only apps that showed differences of greater than 1%. A situation in which a Meaning Accuracy score was higher than a Total Accuracy score would indicate that although there were errors in the transcript, the errors did not affect the meaning of the message. This was not the case in this study, these results suggest that the errors produced by these speech-to-text technologies were not inconsequential (for example, occasional missing morphological markers or spelling mistakes), but were errors that impacted the meaning of the message for the reader.

DISCUSSION

The captioning accuracy for most of the apps and software evaluated in this study was quite good, with 10/12 conditions resulting in >90% accuracy, although none achieved the accuracy of the two gold standards, CART and post production commercial captioning. For a few apps and software, accuracy exceeded 95%. This certainly demonstrates the improvements in today's speech-to-text technologies from its early years, but challenges with accuracy remain, and the ability of the user to impact or improve accuracy is very limited. While some apps and software allow the instructor to attempt to fine tune the accuracy of the speech-to-text captioning (for example, Camtasia Studio allows the instructor to enter new vocabulary with a spoken example so as to "train" the software, and other apps allow the user to manually correct the transcript), this is not feasible for a typical postsecondary instructor and student. The argument for speech-to-text captioning is that it provides an easy-to-use, inexpensive way to provide access. Using technology which relies on training the software, would require each instructor (since different instructors may have different speech patterns) to do this, adding additional costs in instructor time and software costs. Therefore, users of automated speech-to-text captioning are reliant on the developers to continue to improve accuracy. This study suggested that two areas that competent users of English identified as problematic were homophones and punctuation, both of which are difficult for speech-to-text technologies without improvements in the ability of the algorithm to understand the meaning of the message to identify and correct errors.

One key question still remains: when is accuracy 'good enough'? Clearly, the higher the accuracy level, the better, but there seems to be no guidance in the research literature on the definition of an acceptable accuracy level for the average consumer. There is little research on the use of speech-to-text captioning in "real life" conditions under which a consumer might use captioning (such as in a classroom), other than a small body of research from the Liberated Learning Consortium. Much of the research on speech-to-text captioning accuracy in real world applications has focused on broadcast news than on educational use. Leetaru (2019) reported on Google Speech-to-Text transcription for 812 hours of broadcast news (CNN, MSNBC, PBS, NBC, CBS, ABC), and found 92% agreement between Google and (human generated) closed caption transcript. However, speech-to-text accuracy varied, sometimes significantly, by news channel, presumably due to speaker variables and content. Bokhove and Downey (2018) assessed YouTube and diycaptions.com to transcribe interview data and a grade 7 math classroom recording for research purposes. Accuracy ranged from 64 to 69% which the authors described as "decent" as a first step for researchers needing to transcribe video/ audio interviews (assuming that a human researcher would then go through the transcript and correct the errors).

Even the Described and Captioned Media Program, a key online resource for providers and consumers, does not have a guideline for accuracy, other than "errorless captions are the goal". What level of accuracy is required so that administrators, teachers and students could be assured that a particular technology should be implemented? The answer to this question likely differs depending on the situation and the needs of the user (for example, a student who needs captioning as a backup and only glances at it occasionally, versus a student who does not use amplification and relies entirely on captioning for access to instruction). Answering the question of "when is accuracy good enough?", then, requires identifying variables which most interfere with, or facilitate comprehension by, each individual user. What happens in the moment when errors occur (i.e., what strategies does the student have when they encounter comprehension difficulties)? And, perhaps most importantly, how do instructors and users evaluate whether captioning has supported learning or not (i.e., how does one evaluate whether a student has better comprehension with captioning than without it)? The results of this study suggest that some speech-to-text captioning technology may be usable in a classroom (assuming that audio quality can be controlled). In particular, speech-to-text technology created by Otter, Google and Microsoft—companies which have a particular commercial interest in voice recognition for a variety of applications—produced accuracy rates in this study greater than 95%. Given a situation where one can ensure good audio and clear speech by the instructor, with a student who is a strong, adaptive reader, trial of the most accurate speech-to-text technology by students and instructors is worthy of consideration.

Speech-to-text technology is an important advancement in providing access for students who are deaf or hard of hearing but it is important to look at captioning through a language and literacy lens as much as through a technology lens. Since captioning is a representation of spoken language, students must have language competency that allows them to understand the message, and, equally importantly, recognize when they have misunderstood and apply a comprehension repair strategy. A student with poor language or literacy skills may not find captioning to be useful if he/she is not able to read effectively. Jelinek Lewis & Jackson (2001) studied comprehension of captioned television programs by both deaf and hearing students, and found that deaf students' comprehension was lower than hearing students, and that comprehension of questions was significantly correlated with reading competency for deaf students. Captioning is first and foremost access through reading, so student literacy must always be factored into evaluating the efficacy of captioning. An important area for future research would be to investigate the impact of captioning errors on users with a range of literacy skills, as it may be the case that for students with stronger language and literacy skills, lower captioning accuracy rates can be tolerated.

Limitations of the Study

The primary limitation of the study in generalizing results to actual classroom use is the fact that acoustical conditions were intentionally optimized in order to reduce background noise, distance, and reverberation to minimal levels. It will be important to repeat this study in an actual classroom during a live lecture with both the instructor and students in the room. Accuracy rates are predicted to be significantly lower under real world conditions. In addition, while the methodology aimed to recreate a realistic lecture experience, reading the transcript of the lecture may have changed the speech patterns of the lecturer (as compared to speaking spontaneously), which may have had some impact on speech recognition. The lecture evaluated was also quite short. While it seems reasonable to conclude that results for the entire lecture would not differ significantly from the shorter sample, it is possible that the instructor's voice or speech characteristics could change over the course of an entire class (e.g., with vocal fatigue).

Conclusion

The accuracy of speech-to-text technologies evaluated for live lecture conditions in this study was generally very good, consistently greater than 90% and sometimes greater than 95%. The results of this study suggest that some speech-to-text captioning technology may be usable in a classroom. In particular, speech-to-text technology created by Google and Microsoft—companies which have a particular commercial interest in voice recognition for a variety of applications—produced accuracy rates in this study greater than 95%. However, it is important to note that the speech input to the speech-to-text technologies had excellent audio quality, essentially no noise interference, and represented a native English speaker. Effective results in a real postsecondary classroom would depend on the clarity of the speech signal that can be provided. For example, use of a high quality microphone worn by the instructor is likely mandatory for reasonable accuracy; a student simply placing his/her phone on a desk, even in the front row of the class, is unlikely to provide meaningful captions. Given a situation where one can ensure good audio and clear speech by the instructor, with a student who is a strong, adaptive reader, trial of the most accurate speech-to-text technology by students and instructors is worthy of consideration.

Speech-to-text technologies for video captioning were found to be very accurate, increasing the potential for instructors and students to be able to create audio and/or video resources, assignments, and materials, which can easily be made accessible for students who require captioning. Apps for small group work communication were also found to be very accurate, opening up more possibilities for better student interaction (face-to-face or in group chat scenarios). Results of this study suggest that, while use of speech-to-text technology in live lecture situations still poses challenges in ensuring reliable accuracy in a typical postsecondary classroom, educators and students should feel empowered to begin exploring use of this technology in classrooms.

ACKNOWLEDGEMENT

This research study was supported by Minor Research Grant from the Faculty of Education at York University, Toronto, Ontario, Canada

REFERENCES

Aish, S., Jain, U., & Jain, E. (2019). Study of acoustic problems in SDPS Women's College classrooms. *International Research Journal of Engineering and Technology, 6*(9), 1133-1139.

Bain, K., Basson, S. H., & Wald, M. (2002, July). Speech recognition in university classrooms: Liberated learning project. In *Proceedings of the Fifth International ACM Conference on Assistive Technologies* (pp. 192-196).

Block, M. H., & Okrand, M. (1983). Real-time closed-captioned television as an educational tool. *American Annals of the Deaf, 128*(5), 636-641.

Burgstahler, S. (2015). Opening doors or slamming them shut? Online learning practices and students with disabilities. *Social Inclusion, 3*(6).

Cawthon, S. W., Leppo, R., & pepnet 2 Research and Evidence Synthesis Team. (2013). Accommodations quality for students who are d/Deaf or hard of hearing. *American Annals of the Deaf, 158*(4), 438-452.

Consortium for Research in Deaf Education (CRIDE). (2018). *CRIDE report on 2017 survey on educational provision for deaf children.* Retrieved from https://www.ndcs.org.uk/media/4722/cride-2018-uk-wide-report.pdf

Crandell, C. C., Kreisman, B. M., Smaldino, J. J., & Kreisman, N. V. (2004). Room acoustics intervention efficacy measures. *Seminars in Hearing, 25*(2), 201-206.

Fichten, C. S., Asuncion, J. V., Barile, M., Fossey, M., & Simone, C. D. (2000). Access to educational and instructional computer technologies for post-secondary students with disabilities: Lessons from three empirical studies. *Journal of Educational Media, 25*(3), 179-201.

Gernsbacher, M. A. (2015). Video captions benefit everyone. Policy *Insights from the Behavioral and Brain Sciences, 2*(1), 195-202

Glasser, A. T., Kushalnagar, K. R., & Kushalnagar, R. S. (2017, October). Feasibility of Using Speech-to-text with Voices of Deaf and Hard-of-Hearing Individuals. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility* (373-374).

Grasgreen, A. (2013, September). Questionable captions. *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/news/2013/09/27/maryland-lawsuit-over-captioning-deaf-not-unique-likely-wont-be-last

Ibrahim, S.A. (2020). Room mode analysis for classrooms: a case study in the College of Engineering. *IOP Conference Series: Materials Science and Engineering*. 870(1), 107.

Jelinek Lewis, M. S., & Jackson, D. W. (2001). Television literacy: Comprehension of program content using closed captions for the deaf. *Journal of Deaf Studies and Deaf Education, 6*(1), 43-53.

Kawas, S., Karalis, G., Wen, T., & Ladner, R. E. (2016). Improving real-time captioning experiences for deaf and hard of hearing students. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, 15-23.

Kent, M., Ellis, K., Latter, N., & Peaty, G. (2018). The case for captioned lectures in Australian higher education. *TechTrends, 62*(2), 158-165.

Long, G., Vignare, K., Rappold, R. P., & Mallory, J. R. (2007). Access to communication for deaf, hard-of-hearing and ESL students in blended learning courses. *The International Review of Research in Open and Distributed Learning, 8*(3).

Madbouly, A. I., Noaman, A. Y., Ragab, A. H. M., Khedra, A. M., & Fayoumi, A. G. (2016). Assessment model of classroom acoustics criteria for enhancing speech intelligibility and learning quality. *Applied Acoustics, 114*, 147-158.

Millett, P. (2010). Accommodating students with hearing loss in a teacher of the deaf or hard of hearing education program. *Journal of Educational Audiology, 15*(1), 84-90.

Millett, P. (2019). Ontario's high school e-learning still hasn't addressed students with special needs. *The Conversation*. Retrieved from https://theconversation.com/ontarios-high-school-e-learning-still-hasnt-addressed-students-with-special-needs-121612

Millett, P., & Mayer, C. (2010). Integrating onsite and online learning in a teacher of the deaf and hard of hearing education program. *Journal of Online Learning and Technology, 6*(1), 1-10.

Morris, J., & Mueller, J. (2014). Blind and deaf consumer preferences for android and iOS smartphones. In *Inclusive designing* (pp. 69-79). Springer, Cham.

Nguyen, S. (2019, June 25). The best voice to text apps of 2019. [Blog post]. Retrieved from https://www.thebalancesmb.com/best-voice-to-text-apps-4583053

Oberteuffer, J. A. (1995). Commercial applications of speech interface technology: An industry at the threshold. *Proceedings of the National Academy of Sciences, 92*(22), 10007-10010.

O'Callaghan, F. V., Neumann, D. L., Jones, L., & Creed, P. A. (2017). The use of lecture recordings in higher education: A review of institutional, student, and lecturer issues. *Education and Information Technologies, 22*(1), 399-415.

Powell, D., Hyde, M., & Punch, R. (2013). Inclusion in postsecondary institutions with small numbers of deaf and hard-of-hearing students: Highlights and challenges. *Journal of Deaf Studies and Deaf Education, 19*(1), 126-140.

Qiu, L., & Benbasat, I. (2005). Online consumer trust and live help interfaces: The effects of text-to-speech voice and three-dimensional avatars. *International Journal of Human-Computer Interaction, 19*(1), 75-94.

Ranchal, R., Taber-Doughty, T., Guo, Y., Bain, K., Martin, H., Robinson, J. P., & Duerstock, B. S. (2013). Using speech recognition for real-time captioning and lecture transcription in the classroom. *IEEE Transactions on Learning Technologies, 6*(4), 299-311. Retrieved from https://journals-scholarsportal- info.ezproxy.library.yorku.ca/pdf/16155289/v06i0004/435_uatcaltroasr.xml

Reed, S., Antia, S. D., & Kreimeyer, K. H. (2008). Academic status of deaf and hard-of-hearing students in public schools: Student, home, and service facilitators and detractors. *Journal of Deaf Studies and Deaf Education, 13*, 485-502.

Takagi, H., Itoh, T., & Shinkawa, K. (2015, May). Evaluation of real-time captioning by machine recognition with human support. In *Proceedings of the 12th Web for All Conference* (pp. 5).

Tatman, R. (2017). Gender and dialect bias in YouTube's automatic captions. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing* (pp. 53-59).

3Play Media (2020). CVAA Online Video Captioning Requirements and Deadlines [White paper]. Retrieved from https://go.3playmedia.com/hubfs/WP%20PDFs/CVAA-Brief.pdf

Vajpai, J., & Bora, A. (2016). Industrial applications of automatic speech recognition systems. *International Journal of Engineering Research and Applications, 6*(3), 88-95.

VanZant, M. (2005). An analysis of real-time captioning errors: implications for teachers. Thesis. Rochester Institute of Technology. Retrieved from https://pdfs.semanticscholar.org/1c57/1afb01a2119e72e689b928fd0fe6787dc723.pdf

Venkatagiri, H. S. (2002). Speech recognition technology applications in communication disorders. *American Journal of Speech-Language Pathology. 11*(4), 323-332.

Wald, M. (2006). Creating accessible educational multimedia through editing automatic speech recognition captioning in real time. *Interactive Technology and Smart Education, 3*(2), 131-141.

Wald, M. (2008). Captioning multiple speakers using speech recognition to assist disabled people. In *International Conference on Computers for Handicapped Persons* (pp. 617-623).

Wald, M. (2011). Crowdsourcing correction of speech recognition captioning errors. *W4A 2011: 8th International Cross-Disciplinary Conference on Web Accessibility, India.*

Wald, M., & Bain, K. (2008). Universal access to communication and learning: The role of speech-to-text. *Universal Access in the Information Society, 6*(4), 435-447.

Appendix A: Lecture transcript

Lecturer: It's a complicated question, right, about, but it becomes a question at the schools for the deaf about, does, if a child is educated there and they have a cochlear implant or hearing aids or a baha, do they continue to wear them, not continue to wear them, wear them sometimes and not other times, right?

It's a complicated question.

All right.

Let's move on a bit then.

To a kind of related topic, just going to skim through this quickly and then we will take a break in about ten minutes.

So this is a bit of a just a bit of a review of what we talked about before, but we're moving on to this idea of functional hearing assessment, right?

How do you guys figure out whether a child can hear or not?

And it kind of relates to that question of, how do you figure out whether a child is successful with their amplification, whatever it is, because you don't have a test booth and you're not measuring hearing.

But what you're doing out there in terms of functional assessment kind of is the gold standard, right?

Like, what you're doing in terms of observations and checklists and your informal speech perception testing really is just as valid as anything an audiologist does in a sound booth, right?

So don't make this mistake.

A teacher years and years ago had an aided audiogram.

She asked me to come out to see a student.

There we were in the library.

I'm going to do, she said "watch me do a six sound test with this student".

She did hearing aids.

She did.

She did the six sound test.

The kid heard all of them.

He heard, like, an s beautifully.

She said, well here's his aided audiogram and the aided audiogram had thresholds at about 60db, right?

If you had an aided audiogram at 60db at 4,000 hertz, do you think they could hear an s?

No.

Here's the aided audiogram.

I just watched her do a six sound test.

She said "what am I doing wrong?"

Excuse me?

She says "look at the aided audiogram.

This kid can't hear an s."

Okay.

I just saw the kid hear an s.

You saw the kid hear an s.

He is really consistent.

There is no doubt in either of our minds that this kid heard an s.

She was like, "yeah, but."

That's because aided audiogram is wrong, right?

Trust yourself.

Trust your functional hearing assessment because what's on the aided audiogram is just a prediction, right?

Your aided audiogram and your fancy SPLogram is only a prediction.

The way you find out what a kid can hear in the real world is watch what they do in the real world, right?

We're going to cross over a bit into ORCO.

If you have a minute at some point in your life, this is a really actually great TED talk.

It's about how babies develop spoken language.

But we have kind of talked about this idea before babies are hearing before birth, they're developing this kind of template of sounds in their head and then it kind of gets fine tuned to the language that they're hearing most of the time and these random sounds coming in start to turn into words.

Then they start to turn into language, okay?

So for the fulltime students, this should be, like, engraved somewhere, right?

I should be able to say “four conditions of language acquisition” and you should all just go one, two, three, four.

For the parttime students for LLDV next year, you can learn it now.

These are Connie’s four conditions of language acquisition.

You have to have exposure in quality and quantity to accessible language, during meaningful engagement with capable users of the language.

This idea of functional hearing assessment is kind of getting a sense of, are kids able to interact with the world using just their hearing, right?

It’s kind of this idea that Shannon said about being able to use your hearing in a meaningful way.

That’s kind of what we’re talking about.

So especially given that the learning environment gets, like, exponentially more difficult as soon as babies get mobile.

So just a bit of an aside before our break, but it’s an important concept to kind of get your head around.

For the parttime students, the ratio is something we measure when we talk about acoustics, good listening environment, bad listening environment.

Signal to noise ratio, you can measure.

This is a fancy definition.

The signal is what you want to hear, what you want to hear.

And the noise is anything that interferes with what you want to hear, right?

So signal to noise ratio is just the signal minus the noise.

I imagine fulltime students can do this for me.

If we measure the signal and it’s 65 decibels and if we measure the noise and it’s 60 decibels, what’s the signal to noise ratio?

What’s the ratio?

Jodie.

>> 5.

>> yeah, that’s great.

+5 or 5.

>> +5.

>> +5, there you go.

Okay, what if our signal is 55 and our noise is 60?

Then what’s the ratio?

Yes, Caitlin.

>> -5.

>> exactly.

That’s all it’s telling you is how much louder is the signal than the noise?

The bigger the number, the better the listening environment.

As soon as it’s a negative number, it means noise is louder than what you’re trying to listen to and that’s typically not a good thing.

So for the, so for those of you who talked about in your case study about kind of recommendations being to kind of get away from the noise, here’s the problem with that, here’s the problem with the world if you are trying to hear something.

Sound is governed by physics, right?

And you know this.

The farther away you get from the source of the sound, the softer it gets, right?

If you are standing right beside somebody, their voice gets louder.

If you are moving away, their voice gets softer.

This is not something we don’t know.

So what happens to the signal, in which case, in this case, let’s say it’s the teacher’s voice. If you are at the front of the class, the teacher’s voice is louder.

The second row or third row or the back of the room, the teacher’s voice gets softer and softer.

This makes sense to you.

It’s why they do this thing called preferential seating.

Noise doesn’t do that, though.

Noise permeates everything.

It doesn’t matter where you are in the room, the noise is always the same.

Does that make sense?

Because the noise is just this, like, [UNINTELLIGIBLE STUDENT RESPONSE]

I mean, there’s exceptions, right?

Like, if it’s a fan, obviously the fan, close to the fan, the fan is a little bit louder.

That’s a good point, yeah.

So the way sound moves through a room is kind of complicated.

As a general rule, the noise tends to kind of permeate the whole room, right?

So as you are moving away from the teacher, the teacher’s voice is getting softer but the noise isn’t.

The noise is still saying at the same level, right?

So maybe at 2 feet away from the teacher, let’s say, maybe your signal to noise ratio is +10.

The teacher’s voice is 10 decibels louder than the noise.

At 4 feet, maybe it's only, it really does drop this fast.

Maybe it's only +5.

At 8 feet, maybe it's zero.

Kind of at the back of the room with the teacher at the front, suddenly it's minus, okay?

So at the back of the room, the noise is louder than the teacher's voice.

At the front of the room, the teacher's voice is louder than the noise, okay?

You want the rationale for FM systems, that's it, okay?

Because what happens to the signal to noise ratio if you have an FM system on?

Does this apply?

Nope.

With an FM system, it doesn't matter where you are.

You can be anywhere in the room and the signal to noise ratio still holds at a positive number, right?

So it, obviously it's something that, you know, we think about when we think about classroom acoustics.

It's also kind of the idea of why technology doesn't, like hearing aid technology doesn't do a great job at this.

Hearing aid technology talks about noise reduction circuits and all of this fancy stuff you can buy that supposedly gets rid of the background noise.

It kind of, sort of, a little bit… Michael says "no."

Depending on the noise, depending on the signal, maybe a little bit.

It really doesn't work that well.

And part of it is just because acoustics are complicated.

Part of it is also, think about this, if the definition of the signal is what we want to hear, that's not always the person standing in front of us.

Think about this for a classroom, right?

So, for example, if I am in a restaurant with my husband and we're having, because restaurants are noisy, right?

I'm having a conversation with my husband.

What is the signal and what is the noise or who is the signal?

My husband, right?

So my husband is the signal and all of this stuff over here is the noise.

If I hear at the table behind me, I hear somebody said "you know who I ran into the other day, Student 1 Millett," now who is the signal and who is the noise?

Now, this is the signal and my husband is the noise, right?

So it's more complicated than you think, right?

So we talk about FM systems kind of assuming that the teacher is always the signal that you want to listen to and there's a lot more going in a classroom than just the teacher talking.

We try to use passaround mics.

We don't have great technology for that.

We, in our world, we decide there's one signal.

It might not be what the child thinks is the signal.

So let's stop.

Take about a ten minute break.

Then we're going to kind of move on to talk about how are we actually going to measure this ability to function in the real world with your hearing.

Okay?

Appendix B: Small group conversation transcript

Student 1: I guess we might as well get started. Ugh, I hate biology. I can't believe we have to do this stupid project. When is it due again?

Student 2: I'm pretty sure it's the 24th

Student 3: No, it's the 23rd, I looked it up. But we have to hand the powerpoint in to Ms. Pytlovana the day before we do the presentation

Student 1: OK, so here's what we're supposed to do "Name three species (common and scientific name) that have traits that surprised you. What is the advantage of this trait, or do you think it provides a disadvantage? How do you think this trait could have developed, from its ancestors?" What species are we supposed to pick? Does it have to be like mammals or can it be plants or reptiles or something?

Student 2: I don't know, I wasn't even listening when she was talking about it.

Student 3: I vote for wombats, they're cool and I was watching a YouTube video the other day on a wombat and a dog that were friends. And they have square poop so that's pretty surprising.

Student 2: OK, so check this out, I found this video on YouTube "Ten Unique Animals You Won't Believe Exist". There's gotta be something on there.

Student 3: Oooo, I know, colugo, it's like a giant flying squirrel.

Student 2: Can we do bugs? What about this one? The Venezuelan Poodle Moth. It's super creepy too. Oh, never mind, it says it was just discovered a couple of years ago and they don't know much about it so that probably won't work.

Student 1: OK, so do we want to do wombats? What if we do one mammal, one reptile and one insect? Can we do that?

Student 3: Maybe there's more information on Google Classroom. Did anybody look?