Canadian Audiologist

The Official Publication of the Canadian Academy of Audiology

Improving Accessibility with Captioning: An Overview of the Current State of Technology

Published January 22nd, 2019

Pam Millett, PhD, Reg CASLPO

In keeping with the theme of this issue of *Canadian Audiologist*, accessibility, my focus this month is on an accessibility technology that isn't generally a part of our day-to-day practice as audiologists, but which can add tremendous value to what we do to enhance auditory access. This month, I hope to give readers an overview of captioning, as well as some cautions; in many ways, captioning technology is not as advanced as is often assumed. Captioning is not new, most of us are aware of the difference between closed captioning and open captioning, and where to find the setting on our televisions to turn it on. In the past, access to captioning was generally limited to watching TV (if you had the add-on box), and watching the occasional captioned movie in class (for which a teacher of the deaf and hard of hearing had probably scoured the library). With the development of speech recognition technology, this is no longer the case.

We can think of two general types of captioning – real-time and post-production. The primary advantage of post-production captioning is accuracy – a real human has produced the captions and can therefore edit for spelling, meaning and punctuation so that the finished product is polished. Good post-production captioning is the gold standard, as it has the ability to indicate environmental sounds and auditory events (e.g., "door slams"), identify the speaker (e.g., "Professor:"), and provide punctuation for ease of reading. However, post-production captioning can be expensive and time consuming, and there are increased demands for captioning which can be provided on the spot, in real time, for a variety of activities. Web accessibility has become an important topic of conversation across fields which use the Internet for marketing, information and connection (all fields, in other words). Captioning has been shown to support individuals whose first language is not English, to improve their English language skills as well.[1,2] Captioning also allows us to make information available in multiple languages, allowing us to reach a much wider audience. I have even captioned a video in Inuktitut for the Better Hearing in Education for Northern Youth project! You can check it out here.

"Real-Time" Captioning

Real-Time captioning is the technology of choice when captioning is needed on the spot in real time, such as at a conference or a lecture. In the past, real-time captioning always required a human

listener to capture what was said; increasingly, speech recognition software has become able to perform some of the same duties, although accuracy continues to be a challenge in many situations.

Communication Access Real-Time Translation (CART) - Onsite, Remote and Streamed on Television

Many of us have encountered CART at conferences or lectures, or while watching a live news or political broadcast, where captions appears on a screen (or increasingly, on the user's laptop or tablet). CART is really the gold standard of real-time captioning, because it relies on a human listener transcribing the speaker. CART captionists have specialized training and use specific hardware and software to allow them to transcribe phonetically. At York University, for many years, we have used real-time captioning for students but we can now access this service remotely. Through use of a speakerphone and careful consideration of good audio and acoustics, it is possible for a real-time captionist from the West Coast to listen to a lecture in Toronto, and provide real-time captioning which appears on a student's laptop or tablet. This provides seamless accessibility for students, without the potential stigma and distraction of having captions on a large screen. Remote real-time captioning provides increased availability (since the captionist does not have to drive to the location of the event) at a slightly lower cost. Good audio is crucial; however, since the remote captionist typically cannot see the speaker or any audiovisual materials being shown. Real-time captioning is a very effective accommodation for postsecondary education for classroom activities; however, it is important to remember that it does not eliminate the need for notetaking services. I am frequently asked why students with hearing loss might require both real-time captioning and notetaking services, since the transcript of the class is available. A typical transcript for a 2-hour lecture could be up to 40 pages of text; because it is an actual transcript of the class, however, students would still need to read through it and take their own notes which essentially requires the same amount of time (or more) as sitting in the class. Real-time captioning does not serve the same purpose as notetaking, and therefore for best accessibility, both are usually required.

Apps and Software using Speech Recognition

Apps and software that provide "on the go" captioning on a tablet or smartphone have begun to appear, for individuals who have difficulty communicating effectively using spoken language in a particular situation. These apps allow a person with hearing loss to use his/her tablet or smartphone as a microphone and have the speaker's message transcribed into text (for example, talking to a physician, ordering a meal at a fast food restaurant, or having dinner with family members). Ava is a mobile captioning app that can be used on a smartphone with reasonable accuracy for everyday interactions; Interact AS is another technology that requires software on a laptop and a wireless transmitter. There are others emerging on the market with highly variable levels of accuracy – speech recognition will always depend on the sophistication of the underlying algorithm, the clarity of the speaker's voice and distance/noise/reverberation variables, however. Interestingly, there is very little research on the accuracy of these mobile apps and software when used in real life contexts; most seem to go directly from development to market, so being an informed consumer is important.

Increasingly, manufacturers are marketing apps and software using speech recognition for classroom captioning in schools and postsecondary institutions. Using speech recognition captioning as an educational accommodation can be very effective, but introduces many different challenges.[3,4] Captioning in a classroom must be in real-time, with the added challenges of multiple speakers across the school day (for example, for a high school student who has up to 8 different teachers), multiple speakers within one "session" (for example, during a lively classroom discussion with many students contributing ideas), and the use of academic language.[5] Speech recognition algorithms are almost universally based on some type of dictation algorithm, while the vocabulary, grammar, speech patterns, pace and activities of a typical classroom require different underlying algorithms (for example, IBM ViaScribe from the Liberated Learning Consortium, which is based on algorithms specifically developed for academic use).

YouTube, Vimeo, Google Stream and Other Video Sharing Platforms

There is a great deal of video content on the Internet these days for viewers of all ages, and accessing this content can be difficult for individuals with hearing loss. A transcript of the video can be helpful if nothing else is available, but not ideal since it is impossible to watch the video and read a transcript at the same time. Ideally, viewers with hearing loss would like to be able to access captioning for all video content on the Web as soon as it appears. Most viewers of YouTube or Vimeo content will have seen the CC icon on many videos (although not all). The quality of captioning varies widely, from excellent to extremely poor, essentially based on how much human editing was involved. Some videos have captions that were added at the post-production stage by the creator (typically this is the case for videos produced by large corporations or institutions who have the resources to do this). Other videos use the speech recognition feature provided by YouTube or Vimeo. While there has been a significant improvement in the accuracy of this speech recognition technology, it is far from perfect.[6] On YouTube, for example, lighter text indicates words for which the speech recognition algorithm is unsure, and errors are frequent if the speaker has an accent, speaks quickly or if there is noise.

It is important that consumers understand the limitations of speech recognition technology. Poor captioning produced by speech recognition software is in fact worse than no captioning, in my opinion, as it is distracting and requires extra processing time on the part of the reader to identify whether it is in fact an error, to decide whether to ignore it or not. If it is a meaning-laden word or term, the listener must then use precious time and cognitive processing resources to figure out what was actually said.[7] I used an example in a recent presentation of a "real-time" YouTube-provided caption (from a video produced by a university), which showed "anxiety is particularly important because it is one of the Communist mental health problems" instead of what was actually said, which was "anxiety is particularly important because it is one of the commonest mental health problems." Audience members reported being immediately distracted by the error, and noted that it took them a few seconds to get back to paying attention to the actual text (by which time, they had

missed a sentence or two of content).

Other video sharing platforms offer the ability to produce a separate transcript via speech recognition, often with very poor results. For example, Google has recently introduced this service for its Stream video sharing platform, which will provide a transcript of an uploaded video; Bloomfire offers a similar transcription service. My own experience has been that these transcripts can be unrecognizable from the original video, again, depending largely on the quality of the audio. Having the transcript beside the video is also problematic in terms of a user needing to switch his/her visual attention from the video to the transcript and back.

Post-Production Captioning

YouTube and Vimeo

YouTube and Vimeo do offer the ability to add a caption file to an existing video file in post-production so that a creator can make his/her content accessible. There are two ways to do this – the user can upload something called a .srt file, which includes the captions in a time-coded format. Special software (in video editing software such as Camtasia Studio) allows a user to manually add captions to a video file, save it in a .srt format, and upload to YouTube. Captions then appear seamlessly, time coded to the video. YouTube also offers an option to type in your own captions (which are then time coded), or you can use the automatic captioning feature to get started, and then edit for errors.

As professionals whose primary focus is improving communication for individuals with auditory disorders, it is important to have a working knowledge of captioning options. This is important to be able to provide information for clients who may be unaware of the technology (for example, using a mobile captioning app for a work meeting) and to ensure that we are leading the way as role models for accessibility, in advocating for accessibility services as well as ensuring that our own practice demonstrates the gold standard in accessibility (for example, ensuring that video content on our websites is properly captioned).

References

1. Mahdi H. The use of keyword video captioning on vocabulary learning through mobile-assisted language learning. Internat J English Linguist 2017;7(4):1.
2. Rahayu K. Enhancing EFL college students listening comprehension by captioning. ETERNAL (English Teaching Journal) 2018;6(1).
3. Kushalnagar RS, Lasecki WS, and Bigham JP. Accessibility evaluation of classroom captions. ACM Transactions on Accessible Computing (TACCESS) 2014;5(3):7.
4. Shadiev R, Hwang WY, Chen NS, and Huang YM. Review of speech-to-text recognition technology for enhancing learning. J Educat Technol Soc 2014;17(4).
5. Wald M and Bain K. Using speech recognition for real-time captioning of multiple speakers. IEEE MultiMedia 2008;15:56–57.
6. Parton B. Video captions for online courses: Do YouTube's auto-generated captions meet deaf students' needs? J Open Flex Dist Learn 2016;20(1):8–18.
7. Kawas S, Karalis G, Wen T, and Ladner RE. 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 2016;15–23.