www.pipelinepub.com
Volume 19, Issue 2

Advancing Inclusion with AI and Real-Time Captioning

By: Alex Kozlov

Communication Access Real-Time Translation (CART) services transcribe words into text or captions as they are spoken during a classroom lecture, business meeting, public speech, sports event, or arts performance. For people who are deaf or hard-of-hearing, aren't fluent in English, or have auditory processing disorders, CART services are essential to enable basic understanding and participation. In addition, by providing written documentation of an event in real time, CART enhances accuracy and information retention.

Traditionally, CART services have been delivered by highly skilled stenographers. In addition to having to learn to type up to 260 words per minute, a CART transcriber often requires additional specialized training in fields like law or medicine to accurately capture arcane terminology. Because these skills are rare and in demand, CART services tend to be expensive and hard to find.

Today, advances in artificial intelligence (AI) and speech recognition are redefining existing approaches to real-time written translation. By combining human skills with smart software, these innovations offer the potential to significantly expand the availability of CART services. Easier access to CART services, meanwhile, promises to create new opportunities to improve communication and enhance inclusion for people with unique learning styles.

ASR: transcription and context

Automated Speech Recognition (ASR) computer software recognizes speech on two fronts. At the acoustic level, an ASR application "listens" to the sound of a spoken word and produces text corresponding to that sound. While a relatively straightforward task, translating sounds into written words poses a variety of challenges. These include understanding accents, jargon, and

vocal inflections, as well as filtering out background noise. In recent years, significant progress has been made to improve the ability of ASR to understand accents, reduce the need for training an application to recognize an individual's voice, and lessen sensitivity to environmental surroundings.

In addition to recognizing sounds, ASR applications deploy natural language processing models to provide a contextual framework that analyzes the broader meaning of combinations of words. This helps the program, among other things, determine proper spelling and usage. For example, in the statement, "I like my steak medium-rare," the words "medium-rare" provide a context suggesting that the statement relates to "steak" as a food, rather than a "stake" in an organization, or a "stake" driven into the ground. At the same time, the program recognizes that "stake in the ground" refers to a piece of metal rather than a slab of meat. Similarly, contextual analysis can determine that "stake in the ground" is likely an idiomatic expression rather than a literal statement. Based on that determination, the program can more accurately predict the context of the rest of the discussion.

Offline vs. real time

Today's ASR tools are quite adept at transcribing audio recordings after the fact. Readily available and easily affordable tools allow a user to upload an audio file and receive an accurate transcription in as little as five to ten minutes. However, because the transcription is done offline, these tools have the luxury of analyzing the entire discussion before undertaking the transcription. This backwards and forwards perspective allows the program to identify and review the overall context of the discussion, and as a result provide a much more accurate outcome. A CART application, meanwhile, faces the much more difficult task of conducting the contextual analysis on the fly. This means the application must assess the context of each word and sentence as it's spoken, as well as predict the context of the words before they're spoken.

Human in the loop

To address the challenges of real-time speech recognition, researchers are developing end-to-end "transformer" models that apply deep learning techniques to streamline the task of contextualizing words, sentences, and paragraphs. Rather than processing speech on a word-at-a-time basis, these tools analyze entire sentences as clusters. This enables different types of analytics to be applied in parallel, resulting in enhanced speed and accuracy, as well as more robust predictive models.

While significant progress is being made, researchers focused on developing practical applications are not aiming to create automated tools that transcribe speech with 100 percent accuracy. Rather, the goal—and the opportunity—is to supplement AI-enabled ASR with human intervention. This best-of-both-worlds approach allows the intelligent tool to apply processing speed and contextual analysis to transcribe basic words and sentences, while allowing a human interpreter to address nuances, ensure the accuracy of technical terminology, and correct errors.

By leveraging their respective capabilities, an AI-enabled ASR application allows humans to play a role of providing oversight. In some cases, a human will make corrections to the program as it's

working, by interpreting unusual slang, jargon, or technical terms. In others, the human will "revoice," or repeat, unusual or poorly enunciated words in a clear, distinct monotone that the AI can more easily decipher. By executing the bulk of the transcription, meanwhile, the ASR application eliminates the need for specialized training in stenography. As a result, CART services become dramatically more accessible and affordable.

A game-changer for inclusion

From a user's standpoint, access to a real-time transcript of a meeting or discussion creates enormous opportunities for improved learning and inclusion. In a business meeting, for example, CART services allow a deaf person to actively participate and ask questions, and more easily follow the flow of a discussion. While sign language interpreters can enable real-time understanding for deaf participants, sign languages and sign language dialects vary widely, potentially limiting comprehension.

For a deaf person, moreover, access to captioning reinforces information communicated via sign language—much like a hearing person benefits from subtitles when watching a film with heavily accented dialogue. CART services are similarly beneficial to ESL students, particularly those studying medicine, engineering, or other fields with specialized terminology, as well as to individuals experiencing auditory processing disorders.

Looking ahead

Business organizations today are evolving their workplace strategies, aiming to incorporate lessons learned from the COVID-19 pandemic. In many cases, remote work models are playing a significant role. Here, CART capabilities can contribute to effective communication, information sharing, and documentation for teams in disparate locations. For individuals with unique learning requirements working remotely, meanwhile, CART can be a particularly valuable tool in facilitating inclusion and effective collaboration.

Researchers in business and academia continue to explore and unravel the theoretical and practical questions around speech recognition and natural language processing. Emerging use cases include, for example, real-time transcripts of medical procedures and industrial safety protocols to ensure that proper procedures are being followed and to document compliance. Speech recognition programs coupled with AI-enabled sentiment analysis, meanwhile, are providing analytics to help businesses enhance customer experiences.

During the 1980s and 1990s, when researchers were pioneering the development of speech recognition technology, few could have predicted today's ubiquitous presence of Siri and Alexa. As innovation continues, similar surprises may await in our future.