

## **Automated Captioning in the Multiplatform Delivery Age**

*Improvements in speed, accuracy open new doors for captioning live and streamed content*

**Ken Frommert and Eduardo Martinez**

Automation has been infused into innumerable elements of our daily lives. From production and assembly lines to broadcast facilities around the world, the transition to automated processes and workflows now have deep roots, and have forever changed the way we work, shop and entertain.

In broadcasting, the production and transmission of live, manual captioning has long been challenged by high costs, availability, varied latency, and inconsistent accuracy rates. While perfection is impossible due to the speed of live captioning, the transition to more automated, software-defined captioning workflows introduced a new series of challenges.

Closed-captioning is in large part driven by government mandates worldwide to ensure that deaf and hearing-impaired viewers can fully understand and enjoy on-air programming. Closed captions are typically encoded within the video stream and decoded by the TV, set-top box or other viewing/receiving device.

While different mandates on closed-captioning in broadcast television exist around the world, the unifying purpose ensures that deaf and hearing-impaired viewers can fully understand and enjoy the shows they watch. Beyond the hearing impaired, statistics show that one in six viewers worldwide prefer to receive closed captions with their content. This means that as viewers continue to consume content in different ways, technology must evolve to serve changing viewer habits.

### **Deep Neural Benefits**

A common concern across all appliances of automation is the reduction, or outright elimination, of the human element. Closed-captioning is just the latest platform to which these conversations have shifted.

These concerns are beginning to subside as speed and accuracy of speech-to-text conversion continues to improve with the emergence of deep neural network advances. The statistical algorithms associated with these advances, coupled with larger multi-lingual databases to mine, more effectively interpret – and accurately spell out - the speech as it passes through the automated workflow.

Today's strongest automated captioning systems, like ENCO's enCaption4, today approach accuracy rates of 90 percent or higher. The statistical algorithms associated with these advances, coupled with larger multi-lingual databases to mine, more effectively interpret – and accurately spell out - the speech coming through the air feed or mix-minus microphone.

Meanwhile, the faster and more powerful processing of the computing engines within automated captioning technology has significantly reduced the latency to near real-time. This achievement is particularly impressive given that automated captions took between 30-to-60 seconds on many systems as recently as one or two generations ago.

Additionally, as closed captioning software matures, emerging applications to eliminate crosstalk, improve speaker identification and ignore interruptions is improving the overall quality and experience for hearing impaired viewers. Furthermore, the technology is also advancing to support closed-captioning transmission across multiple delivery platforms.

### **New Efficiencies, New Services**

One recent innovation is the introduction of multi-speaker identification, which isolates separate microphone feeds to reduce confusion from cross-talk.

Live talk shows represent an ideal use case. In this scenario, each speaker on the stage is assimilated into the captioning workflow based on their assigned microphone positions, while the software ignores distractions such as low voices and interruptions. The end result is a seamless transition as the conversation shifts between each speaker, eliminating cross-talk and other events detrimental to the viewer experience.

Many of the above improvements are related to recent breakthroughs in machine learning technology for voice recognition. Machine learning not only strengthens accuracy, but it also provides value through detection of different languages and the different ways that people speak.

That intelligence as it relates to different dialects will provide an overall boost to accuracy in closed captioning. Consider a live news operation, where on-premise, automated captioning software now directly integrates with newsroom computer systems with the need for a network connection. This will now help broadcasters strengthen availability – no concerns about a network outage taking the system down – and take advantage of news scripts and rundowns to learn and validate the spelling of local names and terminology.

Automated captioning also enables the applications to be achieved efficiently on a larger scale. The costs are lowered due to the transition from human stenographers to computer automation. And as there is a need to captioning a growing amount of content, there is an economy of scale that drives the cost down even further as broadcasters automate these processes.

As systems grow more reliable and broadcasters grow more comfortable with the technology, they will also find new efficiencies and opportunities along the way. For one, broadcasters that need to cut into a regularly scheduled program with breaking news or weather alerts will no longer forced to find qualified (and expensive) live captioners on short notice.

### **Streaming, the Cloud and Closed-Captioning**

As with many technologies, captioning systems are applicable in both on-premise and cloud configurations. In the latter case, some systems are now offered as SaaS platforms, with monthly fees that include the hardware costs coming out to as low as approximately \$15 per hour for the average rate of use. With stenographer rates sitting at approximately \$150 per hour, this equates to a tenfold savings that can return tens-to-hundreds of thousands of dollars to the broadcaster annually.

However, establishing captioning software in the cloud also extends the service for online audiences outside the local facility, opening the door for efficient delivery of captioned content over streaming networks and delivery platforms. One emerging opportunity for this is the automatic generation of transcriptions for live and archived, pre-recorded content.

As more systems move to software-defined platforms, the captioning workflow for pre-recorded and/or long-form content has been greatly simplified. Post-production staff can essentially drag-and-drop video files into a file-based workflow that extracts the audio track for text conversion. These files can then be delivered in various lengths and formats for a TV broadcast, the web, mobile and other platforms.

This trend aligns especially well for broadcasters and content producers with large volumes of stored media, providing tremendous flexibility to very quickly archive, search, find and recast content tailored to specific audiences and on-demand requests.

Content repurposing software from companies like StreamGuys, previously used for podcasting and specialty broadcast streams, are being tailored for closed-captioning in streaming applications. In this architecture, previously ingested content is recalled through an archived search process. This level of integration also enables users to label and search for specific speakers for improved recognition and tracking, and later search the system for all content related to a program – down to exact, spoken sentences.

With multiplatform reach, broadcasters now have opportunities to caption live and on-demand streams, ensuring that hearing-impaired and multi-lingual audiences watching online are properly served as well. The future of this technology is very exciting, especially with the knowledge that we're really just beginning to reap the fruits of this technology.

*Ken Frommert is president of ENCO, and Eduardo Martinez is Director of Technology at StreamGuys. The two companies will demonstrate their integrated captioning architecture for live and streamed content at InfoComm (Booth 4167).*