Before the
FEDERAL COMMUNICATIONS COMMISSION
Washington, D.C. 20554

	)	
In the Matter of	)	
	)	
Closed Captioning of Video Programming	)	CG Docket No. 05-231
	)	
Telecommunications for the Deaf and Hard of	)	
Hearing, Inc. Petition for Rulemaking	)	
	)	
Telecommunications for the Deaf and Hard of	)	RM-11848
Hearing, Inc. et al. Petition for Declaratory Ruling	)	
And/or Ruling Making on Live Closed Captioning	)	
Quality Metrics and the Use of Automatic Speech	)	
Recognition Technologies	)	
	)	
	)	

COMMENTS OF APPTEK

Mudar Yaghi

CEO
AppTek
1356 Beverly Road, Suite 300
McLean, VA 22101

William B. Wilhelm, Jr.
Christian E. Hoefly, Jr.

Morgan, Lewis & Bockius LLP
1111 Pennsylvania Avenue, N.W.
Washington, DC 20004
Bus.: 202.373.6000
Fax.: 202.739.6001

Counsel to AppTek

Dated: October 15, 2019

TABLE OF CONTENTS

SUMMARY

In response to the Petition for Declaratory Ruling and/or Ruling Making on Live Closed Captioning Quality Metrics and the Use of Automatic Speech Recognition Technologies filed by Telecommunications for the Deaf and Hard of Hearing Inc., *et al.*, AppTek, a leader in automatic speech recognition (ASR) technologies, files these comments encouraging the Federal Communications Commission to consider the application of ASR technology. AppTek supports the evaluation and adoption of ASR technology by broadcasters to improve the accuracy, synchronicity, and completeness of live closed captioning. Currently, the captioning of live programming throughout the United States continues to face significant quality problems and growing consumer complaints.

To address the consumer complaints regarding quality of live program captioning, AppTek believes the Commission should evaluate and, as appropriate, encourage the use of ASR. Recent advancements through machine learning, neural networks, and AI have resulted in ASR technologies being virtually indistinguishable from – and in many cases better than – captioning professionals of live, unscripted programming. For example as discussed below, AppTek's captioning appliance has been rated to have an accuracy rating of 97.86% on the Number, Edition and Recognition Errors (NER) model and has latency between 1.7 seconds and 4 seconds. ASR technology is also available 24/7 and can be utilized on programming when capable live captioning professionals are unavailable.

AppTek further supports a technical examination of ASR by the Commission using any number of neutral benchmarks. While AppTek uses the NER model, any standardized quality metric (Word Error Rate (WER); Automated Caption Error (ACE); etc.) will assist the Commission in determining caption quality across various methods – including ASR. Finally, AppTek encourages the Commission to consider appointing ASR providers to the Disability Advisory Committee (DAC), as the interaction between providers' and the deaf and hard of hearing community will help foster dialog and improvements in the technology.

Before the
FEDERAL COMMUNICATIONS COMMISSION
Washington, D.C. 20554

In the Matter of	)	
Closed Captioning of Video Programming	)	CG Docket No. 05-231
Telecommunications for the Deaf and Hard of Hearing, Inc. Petition for Rulemaking	)	
Telecommunications for the Deaf and Hard of Hearing, Inc. et al. Petition for Declaratory Ruling And/or Ruling Making on Live Closed Captioning Quality Metrics and the Use of Automatic Speech Recognition Technologies	)	RM-11848

COMMENTS OF APPTEK

AppTek, by its undersigned counsel, hereby submits these comments on the Telecommunications for the Deaf and Hard of Hearing Inc., *et al.*, Petition for Declaratory Ruling and/or Ruling Making on Live Closed Captioning Quality Metrics and the Use of Automatic Speech Recognition Technologies.[1]

[1] *Telecommunications for the Deaf and Hard of Hearing, Inc. (TDI), et al.*, Petition for Declaratory Ruling and/or Rulemaking on Live Closed Captioning Quality Metrics and the Use of Automatic Speech Recognition Technologies, CG Docket No. 05-231 (filed July 31, 2019), https://www.fcc.gov/ecfs/filing/10801131063733 ("Petition").

I. INTRODUCTION AND SUMMARY

Founded in 1990, AppTek is a leader in automatic speech recognition (ASR) technologies, as well as, other machine translation technologies. AppTek employs one of the most agile, talented teams of ASR, MT, and NLU PhD scientists and research engineers in the world. Through our advanced research in speech recognition, machine translation, and artificial intelligence, the Company has solved many challenging problems in improving human-quality transcription, language understanding and translation accuracy. AppTek has long-standing affiliations with the world's leading human language technology universities. These affiliations are central to our continuous introduction of new theories and solutions for automating recognition, translation and communication. AppTek's 30 year history of achieving performance goals with our customers across government, global commerce, call centers, telecommunications and media comes from our understanding of their problems and the best application of technology solutions.

AppTek offers a number of solutions including speech-to-text, enterprise translation, subtitling and editing, voice-enabled commerce, live meeting transcription, as well as our live closed-captioning appliance of relevance to this proceeding.[2] Indeed, Petitioners cite AppTek's 2016 FCC briefing and live demonstration of "then-state-of the art ASR captioning solutions specifically designed for broadcast news use" that could significantly improve the display of real-time captioning.[3] This technology has been significantly improved in the three years since that meeting in part because of the Company's work with deaf and hard of hearing community.

[2] *See* Exhibit A; *and see* AppTek, *Live Closed-Captioning Appliance*, https://www.apptek.com/solutions/live-closed-captioning-appliance (last visited October 4, 2019).

[3] Petition at 13; *See also* Letter from William Wilhelm, Counsel for AppTek*,* to Marlene Dortch, Secretary, FCC, CGB Docket No. 05-231 (filed July 29, 2016) (*AppTek Ex Parte*) https://ecfsapi.fcc.gov/file/10729264643292/Apptek%20Notice%20of%20Meeting%20with%20CGB.pdf.

Among its efforts to design ASR solutions that specifically meet the needs of this community AppTek began working with Gallaudet University in 2016 as part of a team working to test and improve technology used to improve closed caption quality.

More than three years have passed since AppTek's FCC meeting and Petitioners aptly note "[i]t is time for a change." Specifically Petitioners: (1) "urge the Commission to…begin in earnest an inquiry aimed at developing objective, technology-neutral metrics for caption quality;" (2) "initiate an inquiry into the state of the art of closed captioning techniques for live television programming and how the varying dimensions of caption quality, including accuracy, synchronicity, and placement affect the accessibility of video programming" followed by a "rulemaking to require live television programming to be captioned at a level that meets or exceeds" designated metrics; and, (3) "urge[s] the Commission to address near-term issues with the use of ASR by" issuing a ruling with near-term guidance on the application of best practices using ASR.[4]

For the reasons below, AppTek supports the adoption of ASR technology by broadcasters to improve the accuracy, synchronicity and completeness of live closed captioning. AppTek believes the Commission should encourage the use of ASR, as appropriate. AppTek further supports a technical examination of ASR by the Commission using any number of neutral benchmarks. Finally, AppTek encourages the Commission to consider appointing any number of ASR providers to membership on the Disability Advisory Committee (DAC). AppTek's technology solutions for the deaf and hard of hearing have greatly improved over time because of its engagement with and feedback from this community.

[4] Petition at iv.

II. CONCERNS ABOUT CAPTION QUALITY SHOULD BE CONSIDERED

The Petition acknowledge that "since the adoption of human-and ENT-centric 'best practices,' Consumer Groups have continued to receive widespread complaints from consumers that quality problems with captions of live programming across a range of markets have continued to persist and even deteriorate in some cases"[5] The Petition refers to four methods to produce captions: stenocaptioner workflows, ENT (electronic newsroom technology), ASR and hybrid workflows. The Petition states that any number of quality issues may appear in each of these workflows. The Petition highlights specific concerns with caption quality:[6]

- Missing Captions for Sports and Weather
- Poor/Bad Accuracy Overall Quality
- Missing Speaker Identification
- Captions Out of Sync
- Missing Captions of Background Noises
- Programs Incompletely Captioned
- Placement Issues
- Significant Omissions and Alterations Impacting Meaning[7]

As the Petition notes and AppTek has observed, omissions are a significant percentage of the error rate in current professional transcription. Live captioning professionals can often become overwhelmed in cases of fast-paced speech in the audio. Live captioning professionals often have to sacrifice completeness to catch up with fast-paced dialogue in a synchronous manner.

[5] Petition at 10.

[6] *Id*. at 10-11.

[7] *Id*. at 13.

As stated in the Petition, "even live human captioners can substantially omit or alter content to a degree that the original communicative intent of the audio track is no longer preserved. Because human captioners excel at editing captions, viewers who are deaf or hard of hearing may not know how much the captions differ from the audio."[8] As AppTek previously observed in 2016, the Company "compared an hour of live programming with captions produced both by the AppTek appliance and human captioning. Because of the nature of live programming and the pace of speech, the Company found that human captioning solutions often contain significant omissions from the actual [dialogue]. In the sample that AppTek reviewed, the company found that if omissions were included, the human captioning provided an accuracy rate of 59% compared with an accuracy rate of 92% using the AppTek solution "out of the box" and without introducing any speaker training."[9] This is exemplified in the video comparing AppTek ASR and live captioning professionals linked in Section III below.

AppTek agrees with the Petitioners "it is clear that quality problems are not restricted to one methodology or technology. Each technology and methodology – human, ENT, ASR, and hybrid modes – demonstrates promise in some contexts but suffers from quality problems in other… ."[10] In a nutshell, one might argue that ENT is perfect as long as everyone sticks to the script and completely fails when not – in which case the user is completely excluded. One might also argue that ASR is more easily and rapidly deployed, always available, and capable of multiple languages, but can make mistakes that live captioning professionals would not. One might also argue that workflows with live captioners are subject to unavailability of staff, errors and omissions that, in some cases, can also alter the meaning of the captioned content.

[8] *Id.* at 12.

[9] *AppTek Ex Parte* at 2.

[10] Petition at 14.

III. THE CASE FOR ASR SYSTEMS IN CLOSED CAPTIONING

ASR systems have clear strengths when it comes to the exact verbatim transcription of speech. Where live captioners have to give in to rephrasing or omission, ASR systems keep up with higher speaking rates. This leads to captioning that is complete and synchronous.

As between live captioning professionals and service providers, there are differences in quality between ASR systems. Certainly, for the deaf community, but also for video programmers and technology providers, it would be supportive to have quality appreciated to avoid, as the Petition states, "a race to the bottom of cost."[11] The sometimes heard claim "of course, humans are better" is easy to make when there is no metric to measure the quality of the output. What we know for sure is that live captioning professionals will remain as good as they are today while the machines are continuously improving. Indeed after the introduction of state-of-the-art technology called deep neural networks, ASR word error rates dropped by half within five years. A modern ASR system can cope with a million words of vocabulary, including vast amounts of names that are unfamiliar to most live captioners. Advances in neural networks and AI are only accelerating, which will lead to further improvements in the technology.

For example, at the time of AppTek's 2016 meeting with the FCC the appliance was capable of captioning 14 languages with upwards of 95% accuracy with training;[12] had a latency of 4 seconds; and, was capable of speaker diarization, speech segmentation, and certain non-speech detection (*e.g.* music, applause). Since then, AppTek has improved accuracy to a

[11] *Id.* at 14.

[12] This accuracy score is based on the Word Error Rate (WER) metric.

97.86% Number, Edition and Recognition Errors model (NER) score and has improved latency to as little as 1.7 seconds.[13]

In addition, to the improvements above AppTek has added additional languages, as well as:

- Punctuation, including periods, commas and question marks
- Capitalization
- Speaker diarization - change detection and formatting
- Custom Glossary - Generate custom lexicons of proper names, characters and dialects for improved accuracy
- Intelligent Word Replacement: Replace words by specific regional dialect to match appropriate spelling. (For example: "Honor" in US English appears as "Honour" Canadian English)
- Smart-Formatting converts dates, times, numbers, currency values, phone numbers and more into more readable conventional forms in final transcripts (US English Only)
- Intelligent Line Segmentation - Split lines at appropriate segments to improve on-screen readability[14]
- Word Confidence - Return confidence level of each word

Many of AppTek's improvements and future developments are facilitated not only by advances in technology and direct feedback from the deaf and hard of hearing community. For example, the speech engine now uses bidirectional long short-term memory (BLSTM) neural networks for robust speech data filtering for acoustic model training and show-specific language modeling. This improves accuracy by understanding context to use phonemes to more accurately anticipate words. Also, it improves the latency and recognition, which impacts

[13] AppTek's appliance outputs raw ASR in 1.7 seconds. In cases of live-to-air for broadcast, post-processing steps provide a total latency of up to 4 seconds.

[14] Intelligent Line Segmentation only applies to subtitling and not live captioning.

accuracy, omissions, and synchronization to dialog. The technological improvements have come through direct outreach to stakeholder groups in the deaf and hard of hearing community. By working closely with users of live captioning, AppTek focused on technological improvements to live captions regarding placement, punctuation, and synchronization that greatly improve the quality of live captions for the reader.

AppTek has further improved captioning performance and robustness since 2016 with significant training and development on media and entertainment data and captioning features. The punctuation models improved with advanced neural networks and additional features (including periods, commas, question marks, capitalization) were added. NLP (Natural Language Processing) features are included to address Intelligent Word Replacement (*e.g.*, region-specific spelling) and smart formatting (*e.g.*, dates, times, currency values, etc.). Speaker diarization and segmentation are enhanced to find the speaker change, speaker grouping, and identifying speech/music/noise/background sounds. Custom dictionary functionality was added to address station or show specific proper nouns. Also, as noted previously, AppTek has been collaborating with Gallaudet University to conduct focus group studies to get feedbacks from hard-of-hearing groups. This feedback helps AppTek identify features and improvements that should be prioritized to better meet the needs of the community.

For the record AppTek provides an example of a comparison of the accuracy of its ASR captioning technology as compared to live captioning professionals at: https://www.apptek.com/post/apptek-outperforms-human-captioners-by-67-video.

An example of broadcast captions generated using AppTek's appliance is provided at: https://www.youtube.com/watch?time_continue=139&v=fAZKOqguRDw.

IV. APPTEK SUPPORTS BENCHMARKING CAPTIONING METHODS

Quality metrics, once defined and established to measure performance and progress, are likely to have a beneficial effect on caption quality. Objective and appropriate quality metrics regularly guarantee performance and progress in many areas – from automobile safety to on-time arrivals. Accordingly, the Commission should consider applying a neutral metric to live professional captioning, ASR, ENT, and other workflows. Quantitative evaluations could provide benchmarks or standards that could be raised over time.

AppTek acknowledges that establishing neutral quality metrics can be difficult and time consuming. There are often disagreements about which model is most appropriate. AppTek notes that the EU and Canada uses a metric called NER.[15] AppTek will frequently test and compare its engine against human generated captions using this technique. In addition AppTek will calculate the simple word error rate (WER) metrics.

AppTek is also aware of other models including one developed by The National Center for Accessible Media at TV station WGBH, as well as Automated Caption Error (ACE) measurements.

Once quality metrics are established it will be easier to determine the benefits and failings of each captioning method. Moreover, as technology improves it will be possible to track these improvements against the metrics. Further, quality standards or best practices could derive out of these tests. As an example, providing lists of proper names, organizations and places (or even scripts) ahead of a broadcast will improve the work of the captioners – and the same holds for ASR – in a measurable way.

[15] *See* David Keeble, *The Canadian NER Trial*, www.nertrial.com (last visited Oct. 4, 2019); s*ee generally* English Broadcasters Group, *Caption Test*, www.captiontest.com (last visited Oct. 4, 2019); Pablo Romero-Fresco & Juan Martinez, *Accuracy Rate in Live Subtitling – NER Model*, http://www.captiontest.com/roehampton%20NER-English.pdf (last visited Oct. 4, 2019).

V. THE FCC SHOULD APPOINT ASR PROVIDERS TO THE DAC

As noted above, two critical factors have paved the way for AppTek's progress in ASR closed captioning technology. First, the company's deep knowledge of speech systems and the underlying technology. Second, the company's engagement with the community of people to whom the captions are provided. While technical knowledge is critical to the development of these systems, it is also critical to understand the end-user and their particular interests.

As a result, AppTek believes it would be useful for the Commission to expand the DAC to include providers of ASR technology. It may even appropriate to run the standard setting and comparison tests at the DAC level. Either way both ASR technology providers and the disabled community would benefit from further engagement with each other. Indeed while this proceeding is limited to closed captioning, AppTek is designing technology that may be of benefit to the community far outside of the context of live television programming. If the FCC can foster increased opportunities for the disabled community and ASR/technology providers to come together, it will improve the adoption of technology to resolve captioning issues. Moreover it will increase the likelihood that those solutions will be designed with the specific needs of the disabled, deaf, and hard of hearing community in mind.

VI. CONCLUSION

For the foregoing reasons, AppTek supports the adoption of ASR technology by broadcasters to improve the accuracy, synchronicity, and completeness of live closed captioning. AppTek believes the Commission should encourage the use of ASR, as appropriate. AppTek further supports a technical examination of ASR by the Commission using any number of neutral benchmarks. Finally, AppTek encourages the Commission to consider appointing any number of ASR providers to membership on the DAC.

Respectfully submitted,

/s/ William B. Wilhelm Jr.
William B. Wilhelm, Jr.
Christian E. Hoefly, Jr.

Morgan, Lewis & Bockius LLP
1111 Pennsylvania Avenue, N.W.
Washington, DC 20004
Bus.: 202.373.6000
Fax.: 202.739.6001

Counsel to AppTek

Mudar Yaghi

CEO
AppTek
1356 Beverly Road, Suite 300
McLean, VA 22101

Dated: October 15, 2019

EXHIBIT A

Live Captioning Appliance

Real-Time On-Premise Same-Language Captioning

www.apptek.com

Overview

AppTek's Live Captioning Appliance is a cost effective, stand-alone server installed with AppTek's automatic speech recognition (ASR) software that delivers fully automated, same-language captions for live broadcast content with accuracy and speed that match or exceed human captioning in real-time.

- Accurate – Upwards of 95% accuracy with training
- Synchronous – Latency as low as 1 second; Average latency of 4 seconds
- Complete - Transcribes all text; No "Dead Air"or missed words.
- Turn-Key - Plug & Play: Connects between media feed and CC Encoder

Apptek's Live CC Appliance resides at the broadcaster's facility and integrates into existing infrastructure and workflows including modern newsroom systems. For stations with a "Digital First" strategy, the Live Captioning Appliance can be used offline to create captions & subtitles for web-stream news content ahead of the scheduled newscast.   Additionally, the appliance can be used to generate metadata for archival solutions and improve value of media or to re-align existing captions and subtitles to synch text with audio. The appliance does not require an internet or telephone connection.

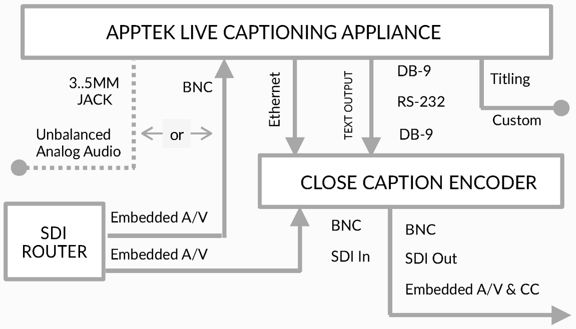
Live CC Appliance Features

- Utilizes Apptek's proprietary Automatic Speech Recognition (ASR) technology to deliver highly accurate, fully automated, real-time, same-language captions for live content
- Pre-trained on broadcast news, sports, weather, and other factual programming
- Custom glossaries and lexicons for improved client-specific accuracy
- Available in multiple languages for same-language captioning
- Speaker recognition and separation with non-speech detection
- Natural Language Processing (NLP) for grammatical breaks and capitalization
- Noise adaptation
- Punctuation and Capitalization Prediction
- Intelligent Line Segmentation
- Profanity Filtering - Censor profanity from displaying on-screen.
- Formatting to denote speaker changes with, ">>"

Additional Features:

- Configurable with AppTek Translate for the creation of subtitles in multiple languages
- Built-in AppTek Workbench Lite for post-editing capabilities

Appliance Connectivity Diagram



Specifications

Supported Languages	Arabic, Chinese, English, French, German, Italian, Korean, Pashto, Persian/Farsi / Dari, Portuguese, Russian, Spanish, Turkish
Supported Interfaces	Option A: SDI embedded audio via BNC Option B: Digital audio via HDMI input Option C: Unbalanced Analog Audio via 3.5MM Jack
Output	Serial Port; Ethernet
Output Formats	Serial and Ethernet output EIA-608 compatible text with compatible Control-A command set. Ethernet output includes plain text and enhanced XML. Export for Workbench Lite or offline transcriptions output in .srt format. includes output
Audio Input Formats	WAV, OGG, FLAC; For offline transcriptions, users can upload a wide array of standard audio file formats. E.g. WAV, MP4, MP3, MPEG, etc.
Management Interface	Browser based user interface
Connectivity Requirements	Standalone on-premise box keeps you in control of your data; No internet or telephone connections required to generate text from speech.
Device Management	Plugs to Internet via Ethernet port to receive software updates
Administration	Administration managed by Apptek, client required to connect box to Internet from time-to-time to receive software updates.

CONTACT US FOR MORE INFORMATION

info@apptek.com | 703-821-5000