Transcribing Air Traffic Control System Command Center Planning Telecons Using Cloud-Based Automatic Speech Recognition

Stephen S. B. Clarke, Aida Sharif Rohani, Jacob Tao and Krishna M. Kalyanam
NASA Ames Research Center, stephen.s.clarke@nasa.gov
NASA Ames Research Center, aida.sharifrohani@nasa.gov
NASA Ames Research Center, jacob.tao@nasa.gov
NASA Ames Research Center, krishna.m.kalyanam@nasa.gov

Your abstract should provide a brief summary of the research conducted, the conclusions reached, and the potential implications of those conclusions. Highlight what is novel in your work. Avoid the use of abbreviations, footnotes, references, or mathematical equations.

Limit your abstract to 750 words.

This paper addresses the challenge of using Automatic Speech Recognition (ASR) technology to transcribe regular teleconferences that happen between FAA Air Traffic Control System Command Center (ATCSCC) planners, stakeholders and air users. These planning teleconferences (aka telecons or planning webinars) are an integral part of managing air traffic in the U.S. National Airspace System (NAS). In particular, the meetings facilitate the creation and modification of various traffic management initiatives (TMIs), that are used to regulate the flow of air traffic. This is typically a human intensive process, requiring specialists to listen to the entire meeting audio (10-20 minutes duration) and inferring the state of the NAS (e.g., weather phenomenon) that was discussed. It would be advantageous to have digital transcripts of the audio and have useful information (e.g., related to TMIs) automatically extracted from the transcripts. In this regard, we are exploring the adoption of state-of-the-art speech to text and Natural Language Processing (NLP) tools that will achieve our objective of digitizing the webinar audio. Unfortunately, the highly technical phraseology present in the audio and limited data availability for model building make ASR difficult. To overcome this challenge, we have taken the critical first step in creating a human transcription dataset from ~20 hours of speech in the ATCSCC audio with the help of subject matter experts.

A novelty of our work is the creation of a ground truth transcription dataset for ATCSCC teleconference webinars, which is particularly important for Aviation domain-specific NLP tasks. Using Microsoft Speech Studio, a cloud-based ASR platform, we have fine-tuned the English pre-trained ASR models (available in speech studio) and achieved an average word error rate (WER) of 6.81%. The baseline ASR also provides a digital version of each planning webinar, making it accessible and text-searchable for future references. Additionally, the transcriptions can serve as a bridge between raw audio data and a range of text-based NLP tasks, such as named entity recognition (NER) and intent classification, potentially enhancing the digital footprint of the webinars and other connected data sources.

Our work has several potential applications. Firstly, the transcriptions can be analyzed to understand the complex decision process of creating, implementing and modifying TMIs and may also contribute to TMI prediction services. Secondly, our dataset and model can be used to develop more accurate ASR systems for aviation-specific language, which can bring about digital communication in the aviation industry (and aid current "voice only" communications,

which are inherently error-prone). Lastly, the transcriptions themselves can be used as a valuable resource for training other NLP models.