

# Enhancing Air Traffic Control Planning with Automatic Speech Recognition

Stephen S. B. Clarke, Jacob Tao, and Krishna M. Kalyanam

*Flight Research Aerospace*, NASA Ames Research Center, [stephen.s.clarke@nasa.gov](mailto:stephen.s.clarke@nasa.gov)

*Universities Space Research Association*, NASA Ames Research Center, [jacob.tao@nasa.gov](mailto:jacob.tao@nasa.gov)

NASA Ames Research Center, [krishna.m.kalyanam@nasa.gov](mailto:krishna.m.kalyanam@nasa.gov)

The decisions made during the Federal Aviation Administration Air Traffic Control System Command Center's planning teleconferences hold significant sway over the National Airspace System. Held every two hours, these teleconferences convene air traffic managers and stakeholders from across the nation to discuss airspace conditions, weather, and constraints, leading to the formulation and adjustment of traffic management initiatives. Given the critical nature of these decisions, the need for accurate and efficient record-keeping is paramount.

In recent years, the application of automatic speech recognition has gained popularity across diverse industries, including aviation. While traditional applications focus on transcribing air traffic control communication, this paper explores a unique application of automatic speech recognition by converting the audio from planning teleconferences into text transcriptions. This innovative approach addresses key challenges in the field, presenting potential benefits for quality assurance, real-time participation, and downstream natural language processing tasks.

A notable breakthrough in the machine learning community, namely the transformer neural network architecture, forms the backbone of the proposed solution in this paper. The transformer architecture's role in this research represents a paradigm shift in the efficiency of automatic speech recognition models. By reducing the amount of in-domain training data required, this architecture allows for the fine-tuning of such models like Whisper, originally pretrained on vast English speech datasets. The adaptability of the transformer architecture proves invaluable in capturing the nuances of aviation terminology and specific language used in planning teleconferences. Leveraging the Whisper model as a baseline, our research details the fine-tuning and validation using a dataset comprising 20 hours of meticulously transcribed planning teleconferences. Notably, the baseline pretrained Whisper model exhibited a word error rate of 18.77%. Through the fine-tuning process, the model achieved a substantial improvement, demonstrating an impressive performance with a reduced word error rate of 6.82%. This substantial decrease in WER not only highlights the effectiveness of the transformer architecture but also emphasizes the practical advancements achieved through the application of automatic speech recognition in this specific domain.

The utilization of automatic speech recognition in planning teleconferences in this work introduces several novelties. Firstly, the creation of text transcriptions offers a valuable tool for quality assurance and facilitates the efficient review of teleconferences. This is an important aspect of the proposed solution, given the time-sensitive and high-stakes nature of decisions made during these meetings. Furthermore, text-searchable transcriptions provide a streamlined approach for locating and validating critical information, potentially saving hours of manual effort in searching through audio recordings. Moreover, our research identifies a key use case for external facilities and stakeholders. In situations where attendance at the planning teleconference

is not feasible, having access to text transcriptions in real-time or shortly after the teleconference ends, proves to be a time-saving and informative resource. This feature enhances collaboration and ensures that stakeholders can stay abreast of important discussions and decisions even in their absence.

Despite the efficiency gains facilitated by the transformer architecture in automatic speech recognition technology, it is essential to acknowledge the human factors in data creation. Subject matter experts play a crucial role in accurately transcribing planning teleconferences due to the specificity and complexity of the information discussed. The research dataset, consisting of 20 hours of transcribed planning teleconferences, forms the foundation for fine-tuning and validating the Whisper model. The achieved word error rate of 6.82% demonstrates promising advancements, particularly in recognizing essential aviation terminology within the teleconferences.

In conclusion, this paper presents a comprehensive exploration of the application of automatic speech recognition in Air Traffic Control System Command Center planning teleconferences, leveraging the transformer architecture for enhanced efficiency. The novel contributions lie in the improved accessibility of decision-making records, real-time participation opportunities for external stakeholders, and the potential for downstream natural language processing advancements. As the aviation industry continues to evolve, the integration of automatic speech recognition technologies holds the promise of revolutionizing decision-making processes and contributing to the overall safety and efficiency of air traffic management.