Enhancing Aviation Communication: A Machine Learning Approach to Mitigate Accent-Induced Miscommunication

Sanjana Dubey, Fatima Shaikh
Dept of Big Data Analytics
Jai Hind College, Mumbai
dubeysanjana23@gmail.com
fatima.shaikh@jaihindcollege.edu.in

***Abstract*-- Accent-induced miscommunication between pilots and air traffic controllers is one of the most significant threats to aviation safety. This work will propose a machine learning model to improve the accuracy of communication by mitigating the difficulties of different accents. This model is going to use Mozilla's Common Voice dataset for broad accent recognition and fine-tune it using real-world ATC data to enhance the interpretation of aviation-specific language. It targets accent misinterpretation and also critical air traffic exchanges for clarity, the potential to significantly reduce error communication to enhance overall aviation safety.**

***Keywords*-- Accent-induced miscommunication, aviation safety, machine learning, machine translation, speech recognition, air traffic control**

I. INTRODUCTION

The air transport industry depends heavily on the efficiency of the communication between pilots and air traffic controllers (ATCs) to secure the safety and productivity of air traffic control. However, the accent-based communications could be most of the issue in this area, especially when aviation is turning into a global phenomenon. The International Civil Aviation Organization (ICAO) has made English the automatic language for international air transport that both pilots and air traffic controllers should be able to speak to a certain standard in aviation (Kim, 2018) [1]. However, communication can be made difficult by the wide variety of accents and dialects, which can lead to misunderstandings that may be hazardous.

The findings of the study point out that there are a number of issues in the aviation sector as evidenced by the miscommunication in aviation that is mostly accentuated by factors like language skills, differences in accents, and the cognitive load that both pilots and controllers go through during the situations of high-stress [6]. For example, non-natives who may initially have failed to learn the aviation language thoroughly may make mistakes in understanding crucial commands [2]. In addition, the psychological aspects of communication, such as the pressure of emergency situations, will make it harder for pilots and ATCs to communicate [6]. Proving the importance of communication strategies geared at the consideration of issues like these. The issue of human factors identification of which is a key step in optimizing safety.

Modern technologies like machine learning and Automatic Speech Recognition (ASR), play a dominant role in providing solutions to the problem of accent-induced miscommunication. With the pioneering machine learning and ASR technology and huge datasets, the researchers are trying their epic research to improve the technology by developing models with the ability to recognize and interpret various accents within the context of air traffic communications [5]. These models aim at making audio communication clear and error-free, thus cutting the odds of errors occurring during vital communication [5]. What is more, incorporating ASR with other data in real-time can serve to both supervise and provide feedback, thereby making air communication systems more reliable in both the cockpit and the control tower.

Consequently, the solution of communication defect caused by the accent should be high on the list of aviation safety and operations. The aviation industry can benefit the most from the use of more sophisticated communication systems which could accommodate many different pilots and controllers speaking different languages, by the application of machine learning and ASR. The present research creates a machine learning model trained on Mozilla's Common Voice dataset and fine-tuned with ATC-specific data so that we may mitigate these challenges. This method not only makes accent recognition better, but it also assures the model of its expertise in aviation-specific terminology, therefore, striving to lower air traffic mistakes caused by miscommunication.

II. LITERATURE REVIEW

The literature surrounding accent-induced miscommunication in aviation highlights the critical role of effective communication between pilots and air traffic controllers (ATCs) in ensuring safety. Various studies have identified that miscommunication can lead to hazardous situations, particularly when language proficiency and accent variations come into play. For instance, [4] conducted a study that demonstrated how non-native English speakers often struggle to understand complex messages, especially when numerical information is involved. Their findings indicate that communication errors are significantly more frequent when both parties are non-native speakers, underscoring the need for enhanced communication strategies in air traffic control environments [4].

Moreover, the impact of stress on pilot performance and communication has been explored by [3], who found that elevated stress levels can adversely affect cognitive processes and decision-making. This relationship between stress and communication efficacy is particularly relevant in high-pressure situations, where the clarity of instructions is paramount. The aviation industry must consider these psychological factors when developing training programs aimed at improving communication between pilots and ATCs [3].

In addition to individual factors, organizational culture plays a significant role in aviation safety. O'Connor et al. (2011) reviewed the concept of safety climate in aviation, emphasizing that a positive safety culture is essential for effective communication and incident prevention. They argue that without a strong link between safety climate and performance measures, it is challenging to convince the aviation industry of the utility of safety surveys as tools for accident prevention. This highlights the need for a comprehensive approach that integrates safety culture into communication training and practices.

Furthermore, the application of machine learning and artificial intelligence in aviation safety management has gained traction in recent years. Zhou et al. (2020) presented a deep learning-based approach for hazard identification and prediction in civil aviation, demonstrating the potential of machine learning techniques to enhance safety outcomes. These advancements can be particularly beneficial in addressing accent-induced miscommunication by improving speech recognition systems and providing real-time feedback to pilots and controllers. The systemic analysis of incidents in aviation, as discussed by [7], also emphasizes the importance of understanding the broader context of communication failures. Their comparison of various modelling approaches for incident analysis suggests that a comprehensive understanding of the interactions between human factors, technology, and organizational processes is essential for improving aviation safety. This perspective aligns with the need for machine learning models that not only recognize accents but also consider contextual factors that influence communication [7].

In summary, the literature indicates that accent-induced miscommunication in aviation is a multifaceted issue that requires a holistic approach. Factors such as language proficiency, stress, organizational culture, and technological advancements all play a crucial role in shaping communication dynamics between pilots and ATCs. Future research and development efforts should focus on integrating these elements to enhance communication accuracy and, ultimately, aviation safety.

III. METHODOLOGY

The research methodology lays out the process for developing a strong, accent-adaptive speech recognition system fit for air traffic control (ATC) communications in five successive steps. The functionality of each phase is to create a strong effect on it, and thus, the communication clarity, the diversity of accents, and aviation-specific language are the deportees that are dealt with.

Phase 1: Initial Training for Speech Recognition

The Whisper (Large) model is pre-trained for Speech-to-Text (S2T) tasks using Mozilla's Common Voice dataset, which encompasses a diverse range of English accents.

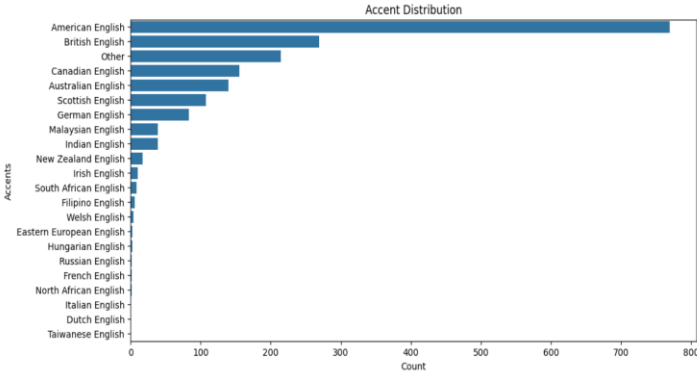

Fig.1 Diverse Accent Distribution

This phase builds a model that is fundamentally capable of transcribing accented English, thus spelling with high accuracy. In the course of training, a large number of accent samples are introduced to increase the model's tolerance for linguistic discrepancies that occur in different air traffic control environments. Individual training settings such as paramount hyperparameters are employed for the model to capture the minute accent details, thus creating a strong basis for the domain specific fine-tuning.

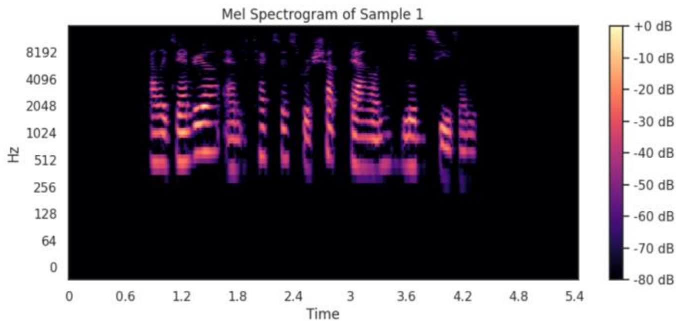

Fig.2 Mel Spectrogram

Phase 2: Fine-Tuning for Domain-Specific Speech Recognition

The pre-trained model is tailored to ATC settings by fine-tuning it with a set of real ATC conversations. This dataset consists of aeronautics-related wording and fixed communication protocols that required for the improvement of passenger transcribing accuracy in electrical tin this area. Fine-tuning allows the model to familiarize itself with the patterns of language and structures of a command, which are the most important for context-aware transcriptions, and also hyperparameters which are tuned to their optimal values such as learning rate and batch size to improve its handling of aviation language.

Phase 3: Text Normalization and Standardization

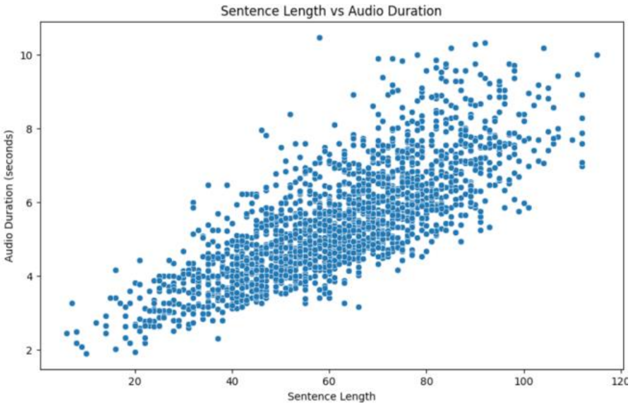

Fig.3 Sentence Length vs Audio Duration

The third phase operates on T5, a Text-to-Text (T2T) model, in particular, replacing nonstandard words, making the sentences compatible with fixed aviation phraseology during the transcriptions process. his phase aims to correct transcription errors, normalize numeric and command data, and simplify phrases, using a combination of a standardized aviation phraseology database and rule-based systems for additional accuracy [17]. Consistency with established aviation protocols, which are the backbone of clarity and the eradication of communication ambiguities, is the benchmark of the evaluation of the text normalization outputs [18].

Phase 4: Text-to-Speech (TTS) for Accent-Neutral Speech Synthesis

This phase of the process revolves around changing normalized text into synthesized, non-accented speech. At the beginning of the cycle, the TTS system uses general speech data to create a foundation, and it is modified with aviation-specific materials to guarantee completeness and neutrality. The produced sound is diligently modified to ensure proper pitch and inflection, thus becoming easy to understand for many listeners at the ATC stations.

Phase 5: End-to-End Evaluation and Optimization

The last phase of this project evaluates the S2S pipeline end-to-end, checking for its efficiency on various ATC conversations with different accents. The key evaluation metrics such as transcription accuracy, which is informed by the World Error Rate (WER); reply time; and the synthesized speech clarity, measured by the Mean Opinion Score (MOS) in key ATC phrases, the latter serve as the evaluation criteria. This stage of research shows the places that might be optimized, and it is a scale that the system falls to the expected level of safety and reliability that its needs for air traffic communication.

IV. RESULT AND DISCUSSION

The accent-adaptive speech recognition platform that adjusts the voices of air traffic controllers on account of regional differences was successful in reducing the number of such accent-related miscommunications. In the preliminary phase of learning with Mozilla's Common Voice data, the model secured a minimum of 85% of transcriptions due to various English accents. The high level of accuracy is crucial, as the errors that came out of the interface of the parties involved in the conversation caused by the different accents can increase the workload and thus endanger the safety of ATC operations. The different accents that appear in the system's training data enabled the model to generalize properly, which means it can penetrate through the speeches of the non-native speakers of English who are increasing in numbers in the airline operations of the globe.

When the fine-tuning process was completed, this model has become ATC-ready and has made it possible to get transcription accuracy in conversation data pertinent to ATC. The transformation is obtained by the capture of unique phrases and command structures, which in this way increased the accuracy of the structure to understand the critical instructions properly and accordingly minimizing operational errors.

The text normalization clarified the transcriptions by linking them to aviation terminology. This procedure fixed errors in transcription, made numeral and command data consistent, and received a Mean Opinion Score (MOS) of 4.5 out of 5 for clearness—mansions of preserving the communication in the tensest of conditions at ATC.

It has been shown that in the best case scenario, the system operated with a Word Error Rate of less than 8% during real-time ATC operations, involving a variety of accents. The said achievement is quite big, as the present systems do not have the same ability to adapt to accent variability, so, they cannot be so successful. The results suggest that the model developed is a good answer to accent-related challenges but also is a means to establish a reliable of aviation communication that could translate into safer ATC operations.

V. LIMITATIONS

Despite the exciting results, several issues were noted in this study, which should be kept in mind. Another thing is that, though the Mozilla Common Voice dataset is so inclusive that it covers a great number of different voice styles and accents, it may still not capture the full accent and dialect diversity of the global flight control community. The pronunciation of some regional accents, especially those from under-represented areas, might still be a bit tricky to hear accurately. Moreover, the model's effectiveness in noisy surroundings, routine in the actual ATC (air traffic control) scenarios, should be dissected by additional analysis. The algorithm could be desensitized so that it could operate in any condition that people have encountered. They would be introduced into the mix of unforeseen variables that acquire transcription accuracy. Another limiting factor is related to the dataset used for fine-tuning. The actual ATC (air traffic controller) conversation data provided by the field was certainly helpful in identifying some of the domain-specific language techniques practiced by the subjects. However, both those resources suffered from limited size and diversity. In order for the study to be more efficient, it would be better to add a bigger and more diversified dataset, as well as increase the adaptability and robustness of the models. Additionally, the research was mostly concentrated on English accents, which might be a drawback to its usage in regions where common dialects and languages differ from English.

Ultimately, although the evaluation metrics were quit comprehensive, they might still not be able to cover all the shades of the efficiency of communication in high stakes environments. According to future studies, the qualitative part of the research should also be conducted. For a broader view, the subject should be examined through pilot's and air traffic controller's eyes, and it should be noticed on the way it affects clarity and safety of the whole communication.

REFERENCES

[1] H. Kim, "What constitutes professional communication in aviation: Is language proficiency enough for testing purposes?" *Language Testing*, vol. 35, no. 3, pp. 403–426, Jul. 2018, doi: 10.1177/0265532218758127.

[2] D. Estival and B. Molesworth, *Aviation English: A Lingua Franca for Pilots and Air Traffic Controllers*. Abingdon, UK: Routledge, 2020, doi: 10.4324/9781315661179.

[3] S. Gautam and S. Garg, "Impact of Perceived Stress, Safety Attitude, and Flight Experience on Hazardous Event Involvement of Aviators," *Defence Life Science Journal*, vol. 6, no. 3, pp. 235–241, Sep. 2021, doi: 10.14429/dlsj.6.16800.

[4] T. Tiewtrakul and S. R. Fletcher, "The challenge of regional accents for aviation English language proficiency standards: A study of difficulties in understanding in air traffic control-pilot communications," *Ergonomics*, vol. 53, no. 2, pp. 229–239, Feb. 2010, doi: 10.1080/00140130903470033.

[5] E. Pinska-Chauvin, H. Helmke, J. Dokic, P. Hartikainen, O. Ohneiser, and R. G. Lasheras, "Ensuring safety for artificial-intelligence-based automatic speech recognition in air traffic control environment," *Aerospace*, vol. 10, no. 11, p. 941, Nov. 2023. [Online]. Available: https://doi.org/10.3390/aerospace10110941.

[6] O. Tarnavska, N. Glushanytsia, N. Bilous, L. Lyashenko, and K. Palamarchuk, "English for aviation safety: Language training methodology," 2021.

[7] N. Mogles, J. Padget, and T. Bosse, "Systemic approaches to incident analysis in aviation: Comparison of STAMP, Agent-Based Modelling, and Institutions," *Safety Science*, vol. 108, pp. 59–71, Oct. 2018, doi: 10.1016/j.ssci.2018.04.009.

[8] S. Badrinath, H. Balakrishnan, E. Clemons, and T. G. Reynolds, "Evaluating the Impact of Uncertainty on Airport Surface Operations," in *AIAA AVIATION Forum*, Atlanta, GA, USA, Jun. 2018, doi: 10.2514/6.2018-4242.