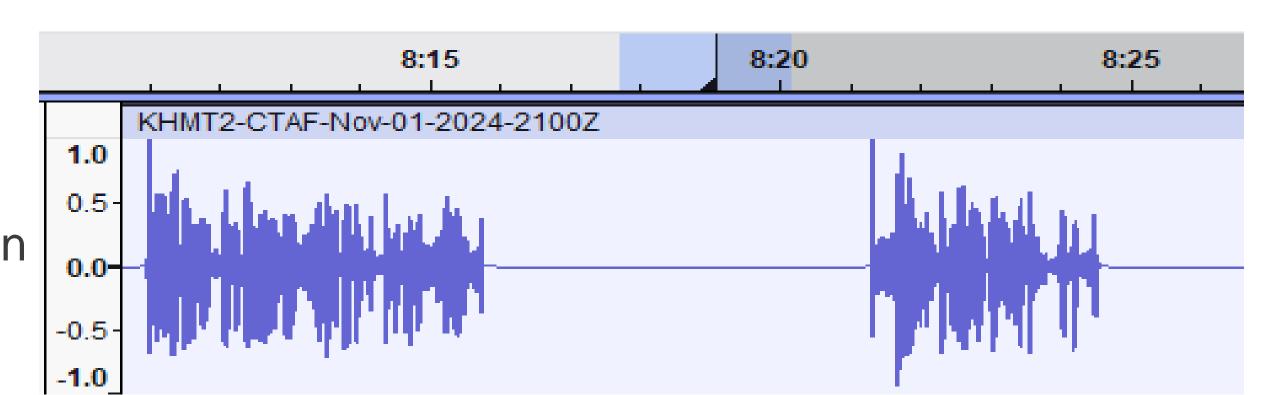


CS 25-346

Al Speech to Text for Military Communications

Team members: Nathan Devore, Nate Eldering, Connor Kohout, Allen Lee | Faculty adviser: Tamer Nadeem, Ph.D. | Sponsor: DoD | Mentor: Clinton | 0.0 |

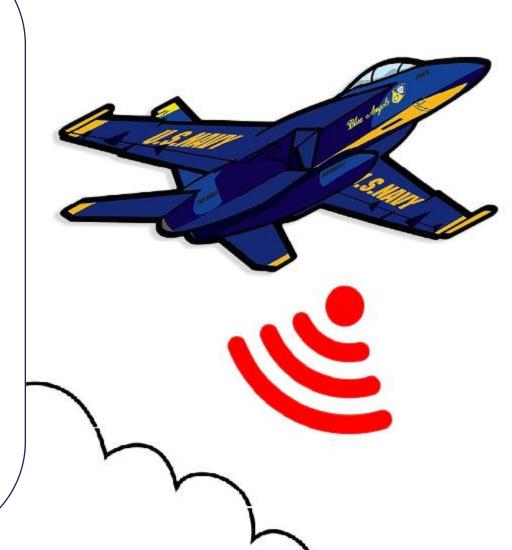


Problem Statement

Military operations occur in noisy environments and involve specific terminology that current speech-to-text models cannot handle effectively. Air communications, in particular, are hindered by wind, radio static, and engine noise, degrading transcription quality.

Initial testing with real air traffic communications shows the Whisper model performs inconsistently in these environments. Our project addresses this challenge by fine-tuning OpenAI's Whisper model, a robust speech-to-text system, to improve transcription accuracy in noisy, mission-critical settings.

Current Design Flow



Audio Collection • Currently, real ATC audio is collected from LiveATC.net o LiveATC has archived audio from air traffic communications Live ATC.net

Text File Data Open AI Whisper Model Output Processing Silence is removed from the • Audio is captioned and Whisper is an open-source output as a .txt file audio captioning tool • The file can then be o Silence can confuse the developed by open-AI and is state of the art whisper model and examined for inaccuracies cause it to hallucinate We have found the o These inaccuracies can • The tool we use is FFmpeg help identify what the medium model transcribes the audio model struggles with so the best during fine-tuning, specific issues can be • The current Whisper model targeted is inconsistent and not very FFmpeg S OpenAl/Whisper

KTOC: 30..2%

K0V4: 76.4%

What is Whisper?

Whisper is an open-source, general-purpose speech recognition model developed by OpenAI, introduced in the December 2022 paper "Robust Speech Recognition via Large-Scale Weak Supervision." Trained on a diverse dataset of multilingual and multi-accented audio, Whisper offers robust performance across various scenarios, excelling at tasks like transcription, translation, and language identification.

It comes in six model sizes, four of which are English-only, each offering a tradeoff between speed and accuracy. This flexibility allows users to select models suited to specific needs, from real-time, lightweight processing to highaccuracy transcription. Whisper's versatility and robust performance make it a powerful tool for a range of speech and language applications.

Whisper's baseline performance can be enhanced through

targeted fine-tuning using aviation-specific data. By having

collecting audio through LiveATC archives, we ensure the

precisely labeled dataset will help the model filter out cockpit

noisy environments. The fine-tuning process will also enhance

and radio interference while improving speech detection in

recognition of proper nouns and the phonetic alphabet.

pilots read scripted communications during flights and

training data matches real-world radio conditions. This

Fine-Tuning

Test Results

To establish a baseline for transcription accuracy, we manually transcribed ATC audio and compared the results to outputs from the Whisper model. Because manually labeling data is so time consuming, we are currently researching better methods. The dataset comprises four audio files of varying clarity, sourced from LiveATC.net archives. The audio was collected from the airports KCHO (Charlottesville Albemarle Airport, Virginia), KHMT (Hemet-Ryan Airport, California), K0V4 (Brookneal-Campbell County

Airport, Virginia), and KTOC (Toccoa Airport, Georgia). This provides a robust foundation for evaluating the Whisper model.

Fine-Tune Targets

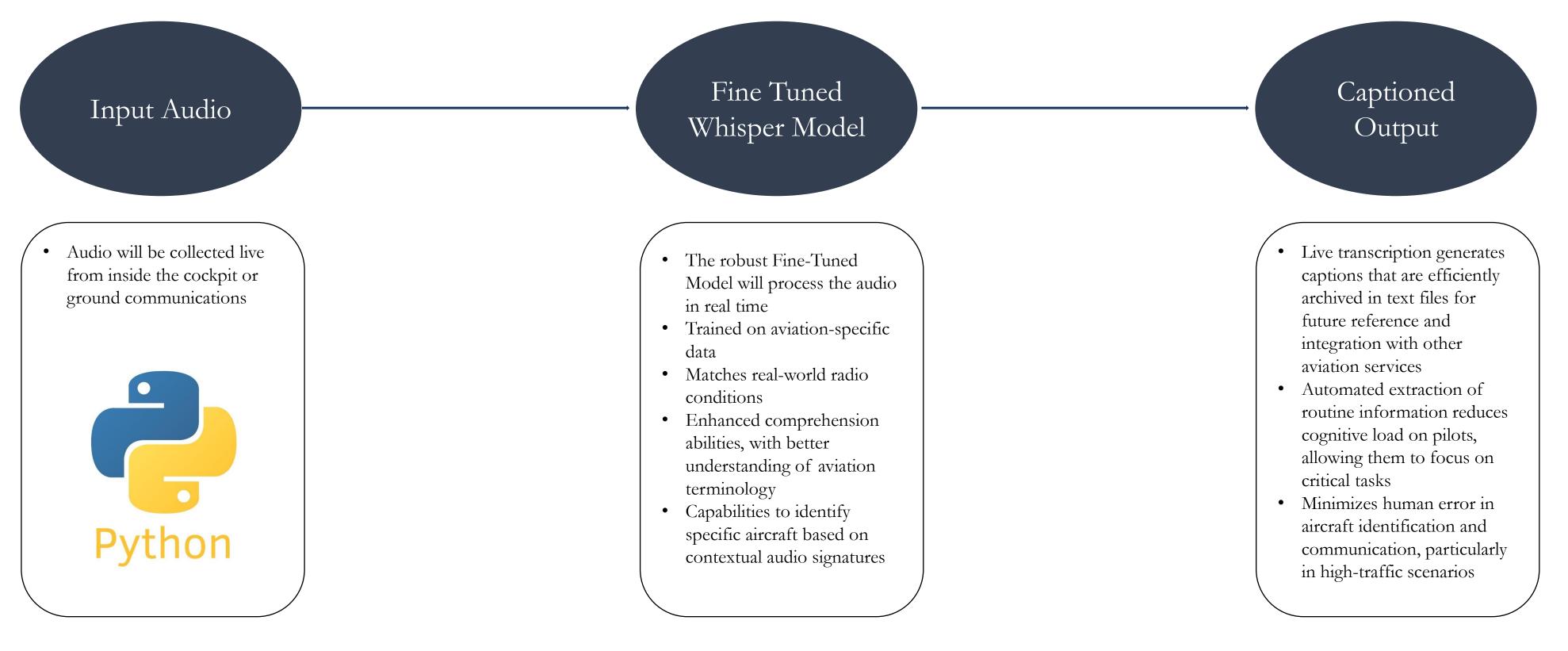
nouns and the phonetic alphabet, significantly hurting its accuracy. For example, it was unable to recognize "Hemet" as "have it." When we fine-tune the model, we will target proper noun recognition and the phonetic alphabet; this includes airport names, company names, aircraft

Ideal Design Flow

KCHO: 84.4%

KHMT: 59.2%

In its current state, Whisper struggles with identifying proper (the name of the airport), incorrectly labeling every instance manufacturers, and types.





College of Engineering