Introduction:
Controlling a drone using voice commands can be an exciting application of technology. In this tutorial, we'll walk through the process of creating a Python script that recognizes voice commands and uses them to control a drone. We'll utilize the Vosk library for speech recognition and assume the use of the Pluto module for drone control.

Libraries:

VOSK

Vosk is a speech recognition toolkit. The best things in Vosk are:

- Supports 20+ languages and dialects - English, Indian English, German, French, Spanish, Portuguese, Chinese, Russian, Turkish, Vietnamese, Italian, Dutch, Catalan, Arabic, Greek, Farsi, Filipino, Ukrainian, Kazakh, Swedish, Japanese, Esperanto, Hindi, Czech, Polish, Uzbek, Korean, Breton, Gujarati. More to come.
- Works offline, even on lightweight devices - Windows,Raspberry Pi, Android, iOS
- Portable per-language models are only 50Mb each, but there are much bigger server models available.

PYAUDIO

PyAudio is a Python library that provides bindings for PortAudio, a cross-platform audio input/output library. It allows Python programs to easily interface with audio devices such as microphones and speakers.With PyAudio, you can easily use Python to play and record audio on a variety of platforms, such as GNU/Linux, Microsoft Windows, and Apple macOS.

Components:

- PyAudio Module: Provides Python bindings for PortAudio, allowing Python programs to interact with audio devices.
- Audio Streams: PyAudio supports the creation of audio streams for recording and playback of audio data.
- Configuration Options: Users can specify various parameters such as sample rate, buffer size, and audio format when creating audio streams.

Install Dependencies

Install the Vosk library for speech recognition.
pip install vosk

Install PyAudio for managing audio streams.
pip install pyaudio

Vosk models:
https://alphacephei.com/vosk/models
Choose the model that best fits your requirements in terms of accuracy, resource constraints, and specialized domain needs.

Usage:

Ensure your microphone is connected and properly configured.
Run the Python script (voice_cmd.py).
Speak the commands near the microphone to control the drone.
For example:
"Start" to arm the drone.
"Fly" to take off.
"Stop" to disarm the drone.
"Land" to land the drone.
You can add or change commands to control the drone.

Testing

DIFFERENT OS

Windows and MacOS:
The voice_cmd.py imports vosk.Model, vosk.KaldiRecognizer, pyaudio, json, and the pluto module or class.
The Vosk model is initialized directly with the model path. PyAudio is also initialized for microphone input, with a larger buffer size.

Ubuntu:
The first snippet imports modules such as os, json, vosk, pyaudio, plutoMultiwii, and Thread. It also imports the pluto class or module.
The Vosk model is initialized with the path to the model and the sample rate. It then initializes a PyAudio stream for microphone input.

MODELS

This document outlines the testing process and results for three Vosk speech recognition models: ***vosk-model-small-en-us-0.15, vosk-model-small-en-in-0.4, and vosk-model-small-hi-0.22.*** The testing aimed to evaluate the performance of these models in transcribing speech in various languages and accents.

Results

1. Vosk-model-small-en-us-0.15
[English (US)]:

Accuracy: Achieved high accuracy in transcribing American English speech.
Speed: Provided real-time or near real-time transcription with satisfactory processing speed.

2. Vosk-model-small-en-in-0.4
[English (Indian)]:
Accuracy: Demonstrated average accuracy in transcribing Indian English speech, accounting for various accents and speech patterns.

3. Vosk-model-small-hi-0.22
[Hindi]:
Accuracy: Showed average performance in transcribing spoken Hindi language with less accuracy.

After testing these models, it is preferable to use an external microphone.Using an external microphone can provide better audio quality and noise isolation compared to the built-in microphone in many laptops or desktop computers
These Vosk Models are intermediate-sized speech recognition models designed to balance accuracy and resource consumption. It offers improved accuracy compared to the Small Model while remaining relatively lightweight.

Further we've tried Google speech to text recognition instead of vosk models.
Here's the difference between these two:

Vosk	Google Speech to Text
Vosk is designed for offline use. It doesn't require an internet connection once the necessary models and packages are downloaded.	Google's Speech Recognition requires an active internet connection as it relies on cloud-based services. This means it may not work in offline environments.
Vosk is providing a real time transcription with a good speed.	Google is taking a little bit more time than vosk to transcript words.
Vosk requires sufficient local processing power and storage space to run the recognition models but doesn't require ongoing network connectivity.	Google's Speech Recognition requires an internet connection and may consume more network bandwidth and battery power due to communication with cloud servers.