STT

1. **Whisper (OpenAI / Whisper.cpp)**

* **Description**: An open-source automatic speech recognition (ASR) model by OpenAI, optimized for offline use via Whisper.cpp. Supports ~60 languages and transcription/translation

REQUIREMENT:  
Python (3.7+)

STEPS:  
pip install openai-whisper torchaudio sounddevice scipy torch  
pip install soundfile

Installation of ffmpeg for tracking the voice instead of providing the audio file

<https://www.gyan.dev/ffmpeg/builds/> -> [ffmpeg-release-essentials.zip](https://www.gyan.dev/ffmpeg/builds/ffmpeg-release-essentials.zip)  
setting up environment variables

base: medium and large are quite close to human-level accuracy on clean speech and even handle accents/language variation well.

**Key Features**:

* High accuracy with low word error rate (WER).
* Supports multiple model sizes (tiny, base, small, medium, large).
* Offline operation on various devices, including Raspberry Pi.

1. **Vosk**

* **Description**: A lightweight, offline STT toolkit with small and large models (100 MB to 1.5 GB). Supports real-time transcription and multiple languages.

**License**: Apache 2.0

* **Key Features**:
  + Runs on low-resource devices like Raspberry Pi, Android, and iOS.
  + Low-latency streaming API.
  + Supports languages like English, Chinese, and more.

Language model based on what model you download from below link.  
<https://alphacephei.com/vosk/models>  
REQUIREMENT:  
pip install vosk sounddevice

We can add real-time transcription (with partial results like "partial": "looking") or language switching via menu.

1. **DeepSpeech (Mozilla)**

* **Description**: An open-source, offline STT engine based on deep learning. Supports Python, C, and JavaScript wrappers.
* **License**: MPL-2.0
* **Key Features**:
  + Low WER (~7.5% for English).
  + Customizable for specific use cases.
  + Note: Development has slowed, with focus shifting to Coqui STT

1. REQUIREMENT:  
   Python 3.6+ installed  
   Install deepspeech Python package  
   Pre-trained DeepSpeech model files (English .pbmm or .tflite, scorer .scorer)  
   A .wav audio file (mono-channel, 16-bit PCM, 16kHz)

<https://github.com/mozilla/DeepSpeech/releases>

pip install deepspeech sounddevice numpy

This is not working for my latest version of python of 3.12 it needs 3.8 version

1. Hugging Face

pip install transformers torchaudio sounddevice scipy

TTS

* PIPER:
* **Python 3.8+** installed.
* Basic knowledge of terminal commands.
* **Git** installed.
* OS: Linux/macOS/Windows (Linux is preferred for simplicity).
* NOT COMPATIBLE BECAUSE OF PYTHON VERSION

**eSpeak**

**On Windows:**

1. Download from: http://espeak.sourceforge.net/download.html
2. Run the installer.
3. Add espeak to your system PATH (if not automatically done).

STEPS:  
 Go to the official download page:  
👉 http://espeak.sourceforge.net/download.html

Scroll to **"Windows"** and download the .exe installer — e.g., espeak-1.48.04-win64-setup.exe.

Install eSpeak. Note the install location — usually:

>>> C:\Program Files\eSpeak\command\_line\espeak.exe

Adding that to the system variables