**PROJECT RESEARCH**

**Abstract:-**

A Project to Fill a Form using Speech Recognition.It leverages speech-to-text technology to allow users to fill out forms by speaking instead of typing.

**Technologies Used:-**

**Backend:-** Flask/Django

**Frontend:-** HTML, CSS, JavaScript

**Speech Recognition:-** Google Speech-to-Text API, Vosk, CMU Sphinx, IBM Watson, Microsoft Azure Speech API

**Database:-** SQLite

**NLP Processing:-** NLTK, SpaCy

**Procedure:-**

**Step 1:** Understand the System Requirements

**Step 2:** Creating a Simple HTML Form

**Step 4:** Set Up the Flask Backend

**Step 5:** Adding Python Speech Recognition

**Step 7:** Storing Data in a Database

**Step 8:** Running the Application

**Best speech-to-text open source models:-**

**1. Whisper ASR**

[**Whisper**](https://www.gladia.io/blog/what-is-openai-whisper) is an open-source speech recognition system from OpenAI, trained on a large and diverse dataset of 680,000 hours of multilingual and multitasking supervised data collected from the web. Whisper can transcribe speech in English and in several other languages, and can also directly translate from several non-English languages into English.

**2. DeepSpeech**

DeepSpeech is an open-source speech recognition system developed by Mozilla in 2017.DeepSpeech uses a deep neural network to convert audio into text, and an N-gram language model to improve the accuracy and fluency of the transcription. Both modules were trained from independent data, to work as a transcriber coupled to a spelling and grammar checker. DeepSpeech can be used for both training and inference, and it supports multiple languages and platforms. Beyond its being multilingual, DeepSpeech runs with the advantage of being quite flexible, and in particular, retrainable.

**3. Wav2vec**

Wav2vec, from the giant Meta, is a toolkit for speech recognition specialized in training with unlabeled data in an attempt to cover as much as possible of the language space covering languages that are poorly represented in the annotated datasets usually employed for supervised training.

**4. Kaldi**

Kaldi is a toolkit for speech recognition written in C++, born out of the idea of having modern and flexible code that is easy to modify and extend. Importantly, the Kaldi toolkit attempts to provide its algorithms in the most generic and modular form possible, to maximize flexibility and reusability (even to other AI-based code outside Kaldi's own scope).

**5. SpeechBrain**

SpeechBrain is an “all-in-one” speech toolkit. This means it’s not just doing ASR, but the whole set of tasks related to conversational AI: speech recognition, speech synthesis, large language models, and other elements required for natural speech-based interaction with a computer or a chatbot.