Report: Emergency Senior Assistant Using Speech Recognition

1. Introduction

In this project, we set out to address a critical issue faced by elderly individuals: their struggle to communicate effectively during emergencies. Existing systems often do not account for the unique challenges that seniors face, such as speech clarity, panic, or difficulty accessing technology quickly. Our goal was to develop a digital assistant that utilizes **speech recognition** and **natural language processing (NLP)** to allow seniors to issue voice commands to call for emergency help.

The system will help elderly users by facilitating communication between them, their caregivers, and emergency services in urgent situations. By leveraging **NLP techniques** such as **speech recognition**, **intent detection**, and **context awareness**, we aim to ensure that seniors can seek assistance promptly when they need it the most. In this week we worked on speech recognition.

2. Approach

2.1. Data Preprocessing

To build this system, we first focused on preparing our data. We worked with audio data representing various emergency scenarios in which seniors might issue commands like "Call for help" or "Emergency." This data was collected from publicly available sources and pre-recorded audio samples. Below are the steps we took to preprocess the data:

- **Noise Reduction**: We cleaned the audio files using libraries like **Librosa** to minimize background noise, ensuring that the speech was clear for the recognition system.
- **Segmentation**: We divided the long audio recordings into manageable segments, focusing on shorter snippets that corresponded to single emergency commands.
- **Normalization**: The volume levels were normalized to ensure that loud and quiet speech samples were treated consistently by the system.
- **Speech Transcription**: We transcribed the audio files using **Google Web Speech API**, converting them into text for easier comparison and analysis during the intent detection phase.
- **Tokenization**: We broke down the transcribed speech into tokens (words) and applied **lemmatization** to standardize the words for further analysis.
- **Stopword Removal**: Common words such as "the" and "is" were removed as they do not contribute to the intent or meaning in the context of emergency commands.

2.2. Real-time Speech Recognition

After preprocessing the data, we implemented a real-time speech recognition system. Initially, we used a microphone input, but for testing and demonstrations, we shifted to using pre-recorded audio files. The system is designed to recognize specific phrases that indicate an emergency and then trigger predefined actions, such as alerting caregivers or contacting emergency services.

3. Speech Recognition from Audio Files

Once our dataset was prepared, we developed the core functionality of the system: recognizing speech from audio files. This is a crucial part of the system as seniors may speak into the system during an emergency. Below is a brief explanation of our implementation:

1. **Loading Audio Files**: We used pre-recorded audio clips to simulate the voice commands that would be issued by seniors during emergencies.
2. **Speech-to-Text Conversion**: By leveraging **Google Web Speech API**, the system converted these audio commands into text, making it easier to process them for intent detection.
3. **Command Processing**: Once the audio was transcribed into text, the system analyzed the detected commands to determine the appropriate action (e.g., sending an alert).

This process is crucial for real-time emergency systems that need to quickly respond to voice commands with minimal delay.

4. Challenges

Throughout this project, we encountered a few challenges, particularly in improving the system's accuracy:

- **Noisy Environments**: Senior citizens may live in environments with background noise, which can interfere with the system's ability to accurately recognize speech. We tackled this by implementing noise reduction techniques.
- **Speech Variability**: Different speech patterns, accents, or panic-induced changes in voice tone can affect recognition. We aimed to train the system to handle these variations by diversifying the training data.
- **Speed of Response**: For emergencies, speed is critical. We focused on optimizing the processing time to ensure that the system responds immediately when a command is issued.

5. Real-World Application

This project has significant real-world applications. By enabling seniors to issue voice commands in emergencies, we provide an extra layer of security for those who might struggle with conventional alert systems. The system can be extended beyond emergency situations and integrated with home automation, health monitoring, or daily assistance features, improving the quality of life for elderly individuals.

---

6. Speech Recognition from Audio Files: Overview

Now that we have a clear understanding of our project's focus, here is a brief summary of the steps involved in **speech recognition from audio files**:

6.1. Audio Preprocessing

We processed pre-recorded audio data to make sure it was suitable for speech recognition:

- **Noise Reduction**: Ensured clean audio for accurate recognition.
- **Segmentation**: Divided audio into smaller, manageable chunks.
- **Normalization**: Ensured consistent volume levels for all audio files.

6.2. Speech Recognition Implementation

We employed **Google Web Speech API** to convert audio into text. The recognized speech was then analyzed to identify the user's intent (i.e., detecting emergency requests):

- **Speech-to-Text**: The system listens to the audio file, converts it into text, and then displays the transcribed text as output.

6.3. Use Case

For example, if a senior's voice command from an audio file was "Help, I need assistance," the system would recognize this as an emergency and issue an appropriate alert. The same process could be applied in real-time through voice commands captured via a microphone.

---

7. Conclusion

In this project, we successfully developed a prototype of an emergency response system for elderly individuals using speech recognition and natural language processing. By preprocessing our audio data and integrating real-time speech recognition, we were able to build a system that could help seniors issue voice commands during emergencies. The potential for such systems to improve elderly care and emergency response is significant, providing greater independence and security to vulnerable populations.

We hereby provide the link for the notebook we worked on

https://colab.research.google.com/drive/16oTXQQMElhmiDGA7hWJeiel9-Z5zYP1Y?usp=sharing