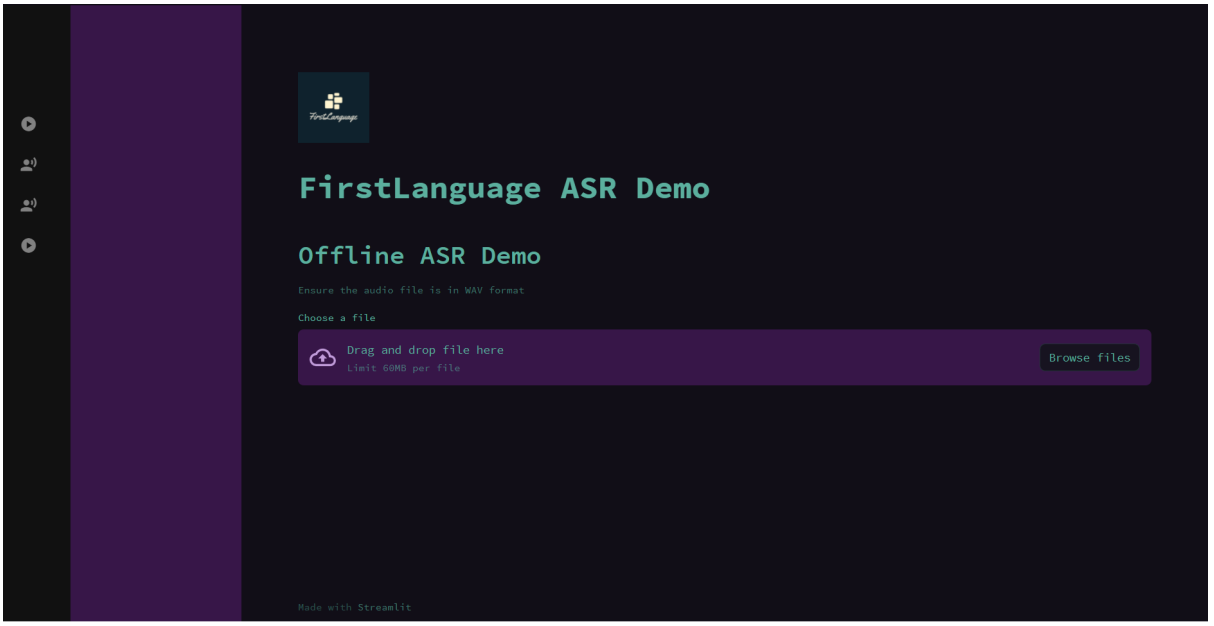
Speech Assistive Web App Documentation

Introduction

The Speech Assistive Web App is an innovative and powerful tool designed to aid individuals with speech impairments in communicating more effectively and efficiently. The app leverages cutting-edge Automatic Speech Recognition (ASR) technology, Text-to-Speech (TTS) capabilities, grammar correction, and next word prediction models to cater to a wide range of speech conditions. By providing accurate transcriptions, generating audio output, and offering real-time streaming ASR, the app aims to empower individuals with dysarthria, aphasia, and stuttering to express themselves with greater confidence and ease.

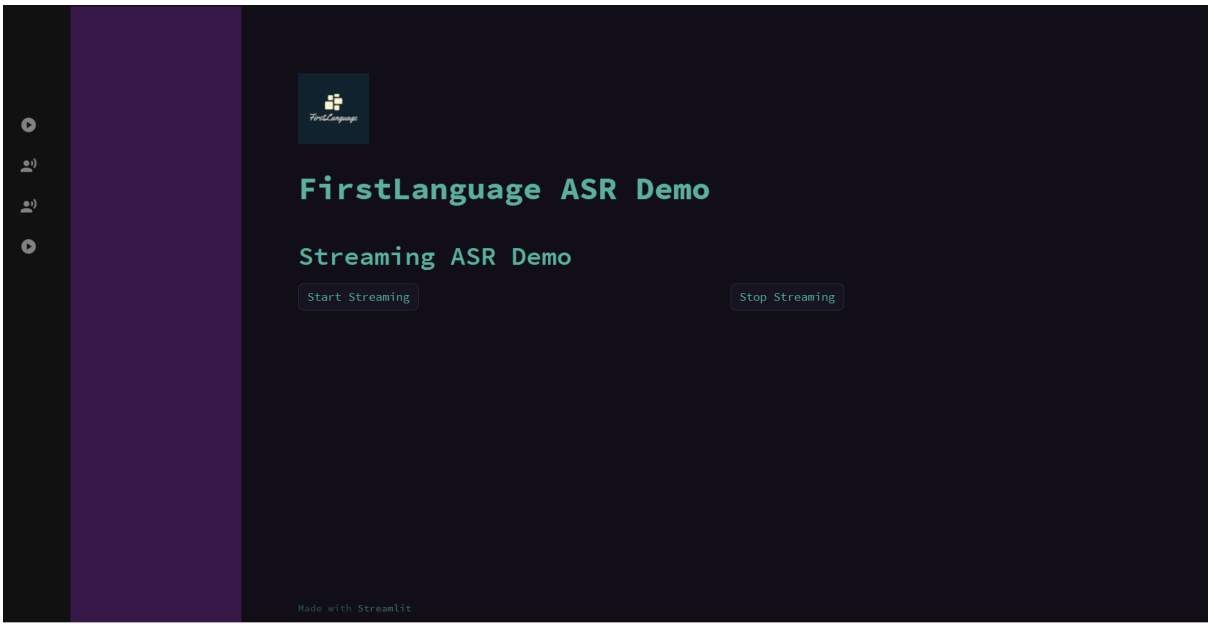
Key Features

1. Offline ASR: The app offers an offline ASR feature that allows users to upload audio files in WAV format for transcription. Users with conditions like dysarthria, where speech muscles are weak, can benefit from this feature as it transcribes their speech with good accuracy, reducing the need for repeated verbal communication. In this feature the user has to upload a prerecorded voice of their own in .wav format and the app will automatically preprocess it and recognise the user's speech. The transcript is then converted into an audio using the TTS, the user can play the audio for whomsoever he may be speaking with.



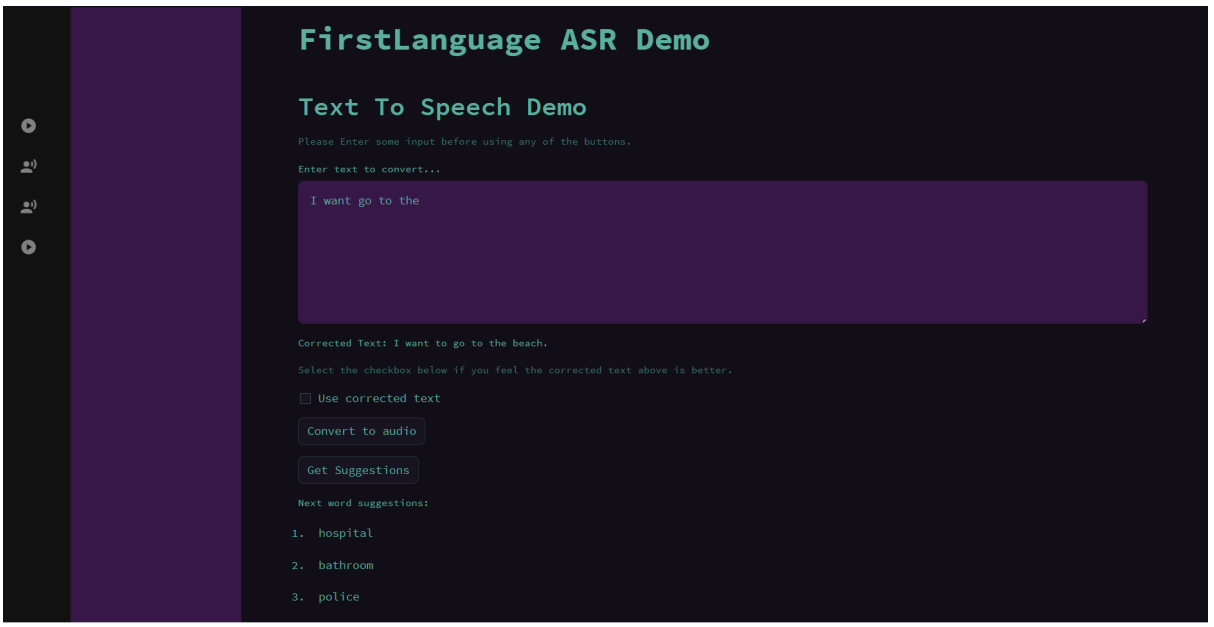
2. Streaming ASR: The app supports real-time streaming ASR, enabling interactive conversations. Users can speak into their microphone, and the app transcribes their speech in real-time, facilitating seamless communication during live interactions. And the user can have the entire conversation transcribed into a single text which is then converted into an audio using the TTS capabilities. The user can have the audio say

anything they said previously out loud.

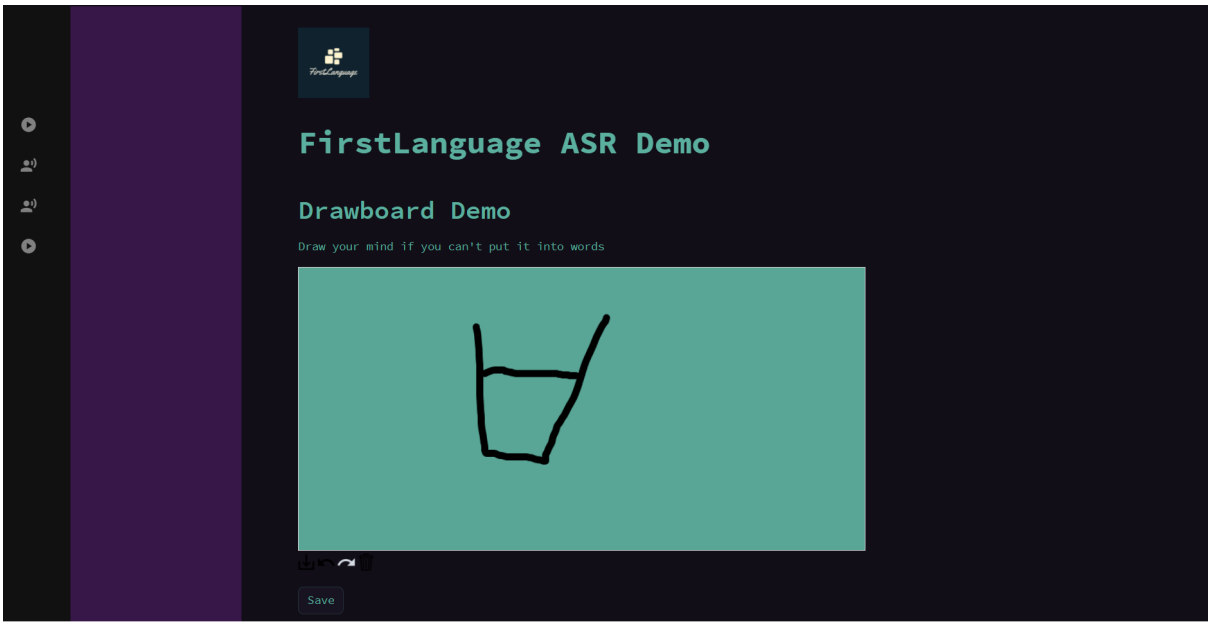


3. Text-to-Speech: Users can input text into the app, and it employs TTS technology to convert the text into audible speech. Additionally, the app incorporates a grammar correction feature powered by Gramformer, ensuring the most accurate text input for optimal TTS results.

    a. Grammar Correction: The Speech Assistive Web App incorporates advanced grammar correction technology powered by Gramformer. By utilising state-of-the-art language models, the app can intelligently identify and rectify grammatical errors in the user's input text. This feature proves invaluable to individuals with aphasia or stuttering, as it helps them convey their thoughts accurately and coherently.
    b. Next Word Prediction: Leveraging state-of-the-art BERT-based models, the app provides next word prediction suggestions. Individuals with aphasia or stuttering can benefit from this feature, as it helps them organise their thoughts and construct sentences more effectively.



4. Drawboard: The Drawboard feature enables users to express themselves visually by drawing images. Additionally, the app offers image captioning, allowing individuals with certain impairments to communicate their visual thoughts.



Use Cases

Dysarthria Use Case

Scenario: A person with dysarthria experiences difficulty speaking clearly due to weak muscles used for speech.

Solution: The Speech Assistive Web App can be utilized to transcribe the person's speech accurately when provided with an audio file. By doing so, it alleviates the burden on listeners to understand slurred or unclear speech, providing a more efficient means of communication.

Aphasia Use Case
Scenario: A person with aphasia has a severely limited vocabulary, making communication challenging.

Solution: The app offers two valuable solutions for individuals with aphasia. Firstly, by speaking into the microphone, the ASR transcribes their speech in real-time, providing text output that is often easier to comprehend than spoken language. Secondly, through TTS, the app can audibly read out the text transcriptions, enhancing the understanding for both the individual and their conversation partner.

Stuttering Use Case
Scenario: A person with stuttering experiences difficulty recalling words or gets stuck on a single word while speaking.

Solution: The Speech Assistive Web App can accurately transcribe all words spoken by the individual, offering an alternative and reliable means of communication. Additionally, individuals can pre-transcribe important content, allowing for smoother and more fluent communication during presentations or conversations.

Articulation Impairment

Scenario: Individuals with articulation difficulties struggle with pronouncing words and sounds accurately. The ASR (Automatic Speech Recognition) component of the app can be a valuable tool for them. As ASR can understand articulated words, it helps in transcribing the spoken input into text. Users can see the recognized text, allowing them to review and self-correct their pronunciation. It provides real-time feedback, which is beneficial during speech therapy sessions and daily practice. The ASR can also be used to track progress over time. By comparing current and previous transcriptions, users, therapists, or educators can monitor improvements and identify areas that still need attention. This data-driven approach enhances personalised treatment plans.

Dyslexia

Scenario: Dyslexia primarily impacts reading and spelling abilities. The ASR component can be beneficial for individuals with dyslexia by converting spoken language into written text. Users can dictate notes, messages, or study materials, and the app will transcribe the content accurately. This helps overcome difficulties in writing, enabling dyslexic users to capture their thoughts without struggling with spelling or handwriting. The app can introduce a "read-aloud" feature where users can listen to the transcribed text being read back to them. This auditory reinforcement can improve reading comprehension and pronunciation.

Testing Results

The Speech Assistive Web App underwent extensive testing across various speech impairment scenarios, yielding impressive results:

Dysarthria Testing Results
Correctly Transcripted: Test cases 002, 004, 008, 009 (Slurred due to painkillers)
Misinterpreted Few Words: Test case 003
All Wrong/Only Few Words Caught: Test cases 001, 010

Aphasia Testing Results
Transcripted with High Accuracy: Test cases 005, 006, 007 (Counting numbers)

Stuttering Testing Results
Perfect Transcriptions: All words spoken in test cases 011, 012, 013, and 014 were accurately transcribed.

Libraries and Models Used

These libraries, models, and APIs collectively form the backbone of the Speech Assistive Web App, offering an array of features to enhance communication and assist individuals with aphasia, stuttering, and other speech impairments in expressing themselves effectively and independently.

1. **Streamlit**: Streamlit is a Python library that enables the creation of interactive and customizable web applications for data visualisation, machine learning, and more. It forms the foundation of the Speech Assistive Web App's user interface, allowing users to draw, transcribe, and interact with various functionalities seamlessly.
2. **PyDub**: PyDub is a Python library for audio processing. In the app, PyDub is used for converting and processing audio files, facilitating the conversion of audio inputs to the required format for offline Automatic Speech Recognition (ASR) testing.
3. **imageio**: The "imageio" library facilitates reading and writing image data, allowing the app to work with images in various formats.
4. **Pathlib**: The "pathlib" module provides a convenient way to work with file system paths and file operations.
5. **Transformers**: The Transformers library, developed by Hugging Face, provides access to state-of-the-art natural language processing (NLP) models, including BERT (Bidirectional Encoder Representations from Transformers). These models underpin the app's grammar correction and image captioning functionalities.
6. **BART Model (Bidirectional and Auto-Regressive Transformers)**:The BART model, from the "transformers" library, is utilised for various natural language

processing tasks, including grammar correction. BART is a denoising autoencoder that combines bidirectional and auto-regressive training to understand and generate human-like text. We have used this model for getting suggestions for what the next Input might be.

7. **Gramformer**: Gramformer is a library that enhances the grammatical correctness of text using transformer-based models. It is integrated into the app's grammar correction feature to identify and suggest corrections for grammatical errors in the user's input text.
8. **The Riva ASR** (Automatic Speech Recognition) and TTS (Text-to-Speech) APIs which are the APIs of FirstLanguage itself are employed in the "Streaming ASR" and "Text-to-Speech" features, respectively. The ASR API enables real-time streaming speech recognition, transcribing spoken words into text, while the TTS API synthesises text into natural-sounding speech, enabling users to listen to their transcriptions.

Future Scope

As we look to enhance the capabilities of our Speech Assistive web app, we have identified several promising areas for future development. These improvements will not only expand the app's usability but also contribute to a more comprehensive and user-friendly experience for individuals with speech impairments. Here are some of the exciting future prospects we envision:

- **Image Captioning and Object Recognition**: We plan to integrate advanced computer vision techniques to enable our app to provide image captions and recognize objects in the drawings made on the drawboard. This feature will assist individuals with visual impairments in understanding the content of images and drawings.
- **History of transcriptions:** We can add a login system to store the transcriptions of a particular users so that they can compare with past transcriptions to find the improvement level of the user via the therapy.
- **Automated File Conversion for ASR**: To streamline the user experience, we aim to implement automated file conversion for audio files used in ASR. This enhancement will allow users to directly upload audio files in various formats, and the app will automatically convert them to the required format, eliminating the need for manual conversions.
- **Incorporating Educational Games**: To further support children with speech impairments, we plan to introduce interactive and engaging educational games. Some game ideas include Word Association Game, Sentence Formation Challenge, Speech Sound Identification, Vocabulary Builder Quiz, Story Completion Game, and Phonics Matching Game. These games will serve as both therapy tools and fun learning activities, aiding in brain development and language skills improvement.
- **Extended ASR Use Cases**: We aim to expand the ASR capabilities to accommodate a wider range of use cases. By fine-tuning the ASR models, we can cater to different

speech impairment types, including dysarthria, aphasia, and stuttering, providing more accurate and personalised results.

- **Optimization for Faster Processing**: To improve the app's efficiency, we will work on optimising the code structure and algorithms. By reducing processing time, users can benefit from quicker responses and a smoother experience.
- **Interactive User Interface**: We plan to revamp the user interface to be more interactive and intuitive. A user-friendly interface will enhance accessibility and usability for individuals with varying degrees of computer literacy.
- **Multilingual Support**: To cater to a broader user base, we aim to incorporate multilingual support for both ASR and TTS functionalities. This will enable users to interact with the app in their native language, making it more inclusive.
- **Integration of More NLP Models**: We will explore the integration of additional state-of-the-art NLP models to provide advanced grammar correction and context-based word suggestions, making the app's communication assistance more robust and accurate.

With these future developments, we are committed to continuously enhancing our Speech Assistive web app, ensuring it remains at the forefront of technology and remains a valuable tool for individuals with speech impairments. As we progress, we will keep our focus on empathy, inclusivity, and technological excellence. Our ultimate goal is to empower and uplift the lives of those with speech difficulties, enabling them to communicate effectively and with confidence.

Conclusion

The Speech Assistive Web App stands as a transformative tool for individuals with speech impairments, significantly improving their ability to communicate effectively. By integrating ASR, TTS, grammar correction, and next word prediction models, the app empowers users to express themselves with greater confidence and clarity. As a result, the app has the potential to enhance the quality of life for individuals with various speech conditions, fostering more inclusive and meaningful communication experiences.