JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

D. Y. PATIL TECHNICAL CAMPUS FACULTY OF ENGINEERING AND FACULTY OF MANAGEMENT, TALSANDE (POLYTECHNIC)

Speech to Text Conversion

Prajakta Vibhute[1], Anuradha Shid[2], Nandini Patil[3], Shreya Gangdhar[4], Jyoti Gurav[5]

Department of computer engineering

ABSTRACT

Speech-to-text (STT) conversion, also known as automatic speech recognition (ASR), refers to the process of converting spoken language into written text. This technology has gained significant attention in recent years due to its potential applications across various domains, including healthcare, customer service, education, and accessibility. The primary goal of STT systems is to accurately transcribe human speech in real-time or from recorded audio data while handling variations in accents, speech patterns, background noise, and other real-world challenges.

The STT conversion process typically involves several stages, including audio signal preprocessing, feature extraction, pattern recognition, and text generation. Initially, the input audio is captured using microphones or recording devices, followed by the extraction of relevant features such as Mel-frequency cepstral coefficients (MFCCs) to represent the speech signal. These features are then processed using machine learning algorithms—ranging from traditional hidden Markov models (HMMs) to more advanced deep learning models such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), including Long Short-Term Memory (LSTM) networks. These models are trained on large datasets to recognize and predict phonetic, linguistic, and contextual elements of speech.

Key challenges faced by speech-to-text systems include handling homophones, speech disfluencies, speaker variability, noise interference, and real-time processing demands. To improve accuracy and

robustness, STT systems often incorporate techniques such as noise filtering, speaker adaptation, language modeling, and context-based corrections.

I. INTRODUCTION

Project presents a novel speech-to-text technology with an integrated PDF conversion capability, all with the goal of supporting members of the handicraft community. By offering a simple tool that converts oral instructions or thoughts into written text, this technology seeks to empower artists and crafts people by removing any obstacles that may arise from using more conventional written approaches. Its capacity to smoothly transcribing text into a standardized PDF format adds to the system's adaptability by enabling effective recording of project details, design concepts, and creative activities. This project, which is specifically designed to meet the demands of handicraft practitioners, aims to improve communication and accessibility for artisans in order to create a more productive and inclusive atmosphere for their artistic pursuits.

The project's speech-to-text technology and built-in PDF conversion capability provide a productive and time-saving way to write letters and other written documents. The task of manually entering text is expedited by allowing users to express their idea orally and having the system accurately copy them. This minimizes the possibility of errors related to human data entering while also saving time. An additional level of convenience is added by having the transcription text in a PDF format. The whole letter generation process is made more efficient by users' ability to swiftly produce documents with a professional appearance without requiring considerable formatting or editing. This feature is especially helpful for people who might be time-constrained or who would rather use a more fluid and natural method of exchange of ideas. Overall, the project helps save time and streamlines the letter writing process, providing a practical and efficient solution for people Who prioritize productivity while creating documents.

II. RELATED WORK

➢ Programming Language: JavaScript

JavaScript is a high-level, dynamic, and interpreted programming language that is primarily used for client-side scripting on the web. For speech-to-text conversion, JavaScript is a great choice due to its access to the Web Speech API, which allows developers to create web applications that can recognize and synthesize speech. The Web Speech API provides two main interfaces for speech-to-text conversion: Speech Recognition and Speech Synthesis. The Speech Recognition interface allows developers to create speech recognition objects that can recognize spoken words and phrases. The Speech Synthesis interface allows developers to create speech synthesis objects that can synthesize text into speech. JavaScript events, such as on result and on error, can be used to handle speech recognition and synthesis events. Popular JavaScript libraries for speech-to-text conversion include Google Cloud Speech-to-Text, Microsoft Azure Speech

Services, and IBM Watson Speech to Text. However, there are also challenges and limitations to consider when using JavaScript for speech-to-text conversion. For example, not all browsers support the Web Speech API, which can limit the compatibility of speech-to-text applications. Additionally, speech recognition accuracy can vary depending on the quality of the audio input, the complexity of the spoken language, and the capabilities of the speech recognition engine.

➢ Frontend: React.js

React.js is a popular JavaScript library for building user interfaces, including speech-to-text applications. Developers can leverage the Web Speech API, React-Speech-Recognition, and cloud-based APIs like Google Cloud Speech-to-Text to build high-quality speech-to-text applications. These libraries and APIs provide simple access to speech recognition capabilities, support multiple languages, and offer real-time speech recognition. When building speech-to-text applications with React.js, developers must consider browser support, audio quality, and language support. Techniques like feature detection and audio processing can overcome challenges like browser compatibility and poor audio quality. Libraries like js PDF enable developers to generate PDF documents from transcribed text, while libraries like Speech Recognition provide real-time speech recognition capabilities. Overall, React.js is a powerful tool for building speech-to-text applications, and its integration with various speech recognition APIs and libraries makes it an ideal choice for developers. This enables developers to build applications that support a wide range of users and provide a seamless user experience.

➢ Backend: Node.js

Node.js is a popular JavaScript runtime environment that provides a robust platform for building high-performance speech-to-text conversion applications. Its fast execution, scalability, and extensive libraries make it ideal for real-time applications. The Node.js ecosystem offers various libraries, including Google Cloud Speech-to-Text and IBM Watson Speech to Text, which provide access to cloud-based speech-to-text APIs. Node.js frameworks like Express.js and Koa.js enable developers to build complex web applications, handling HTTP requests, interacting with databases, and rendering templates. The benefits of using Node.js include fast and efficient execution, scalability, and support for multiple languages.

➢ Server Framework: Express.js

Express.js is a popular Node.js framework for building web applications, including speech-to-text conversion applications. It provides a flexible and modular way to handle HTTP requests, interact with

databases, and render templates. Express.js can be used to create a RESTful API for speech-to-text conversion, allowing users to upload audio files and view transcribed text. Its middleware architecture and routing mechanism enable developers to add support for multiple speech-to-text engines, handle errors, and implement authentication. Express.js can also be used with other libraries to build real-time speech-to-text conversion applications, providing an interactive user experience. By leveraging Express.js features, developers can build high-quality speech-to-text conversion applications that meet user needs.

III. SYSTEM ARCHITECTURE

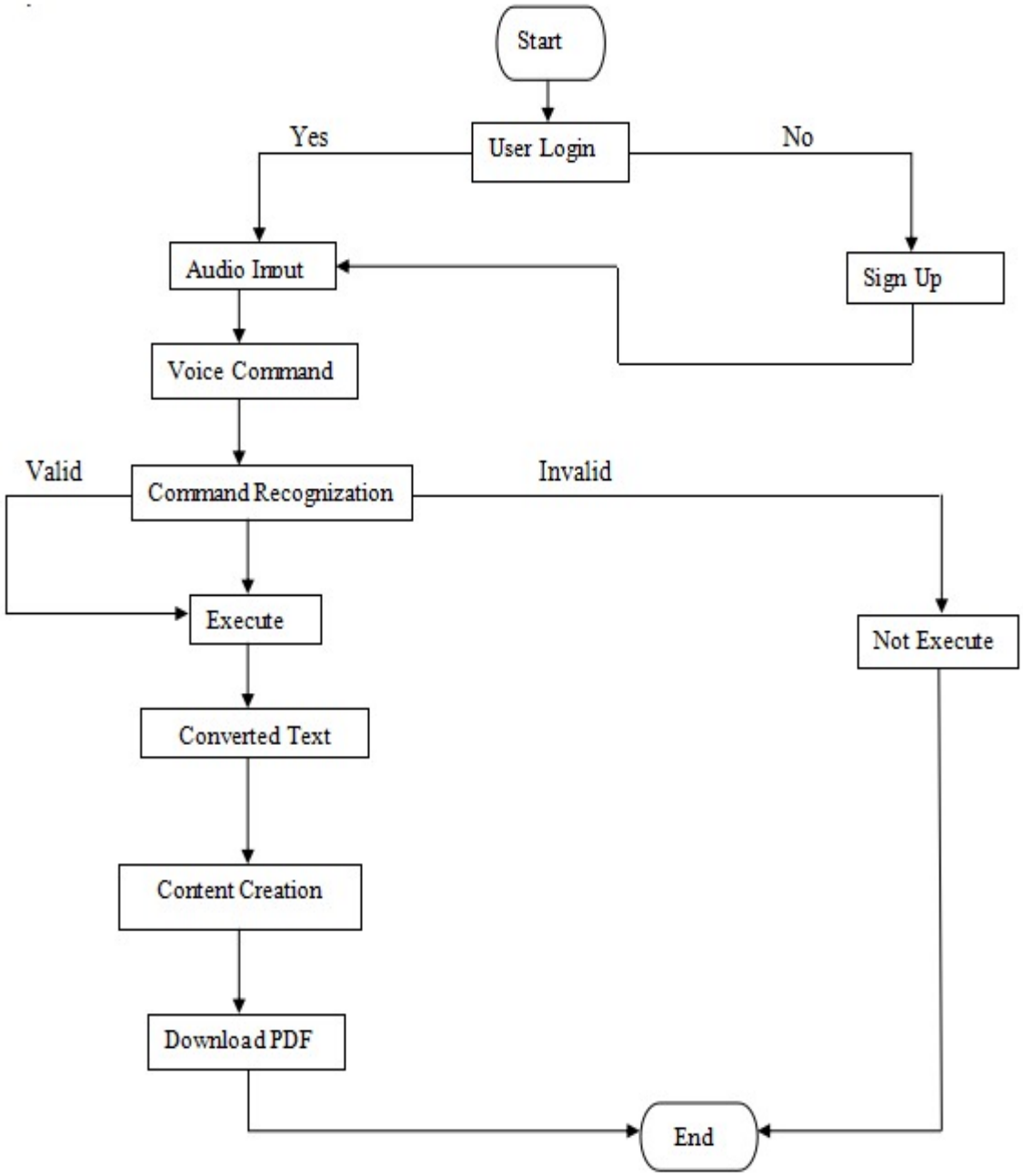

Fig: Proposed System Architecture

IV. IMPLEMENTATION AND MODULES

Module 1: User Module

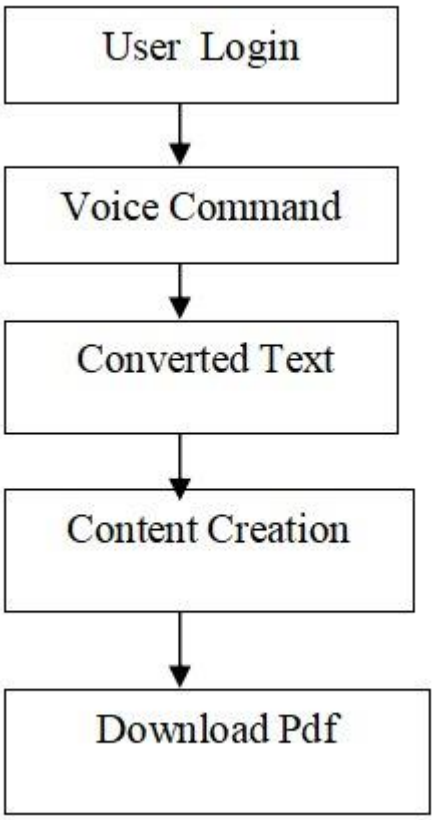

Fig 4.1: User Module

The user module of speech-to-text conversion refers to the component of a system or application that allows users to input their speech, which is then converted into text. In many cases, this process involves a speech recognition engine or module that captures audio from the user's microphone and translates it into written form. This can be implemented on various platforms using different libraries or services.

Module 2: Contain Creation Module

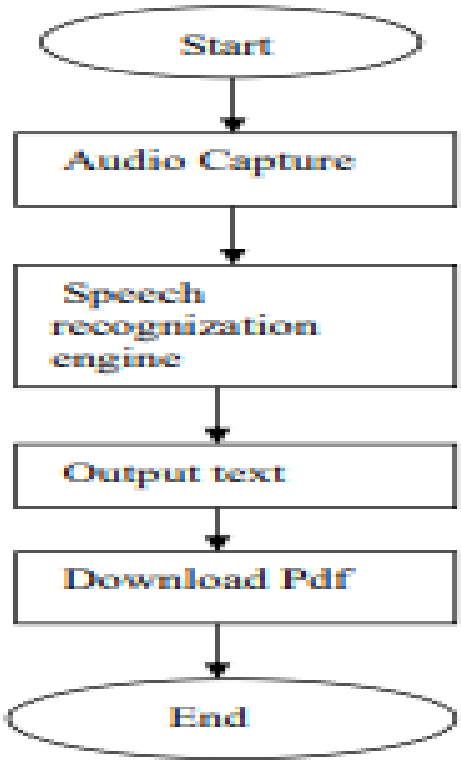

Fig 4.2: Contain Creation Module

The Content Creation Module in the context of speech-to-text conversion is the part of an application or system responsible for creating, managing, and handling the content (i.e., the transcribed text) that results from the speech recognition process. It typically interacts with the output from the speech recognition engine (which converts audio to text) and helps to manage, store, format, or further process that content.

This module plays an essential role in applications where speech-to-text is used to generate, edit, or publish textual content. It can be found in various real-world applications, such as transcription services, voice assistants, real-time captioning, and content creation tools.

Module 3: Voice Command Module

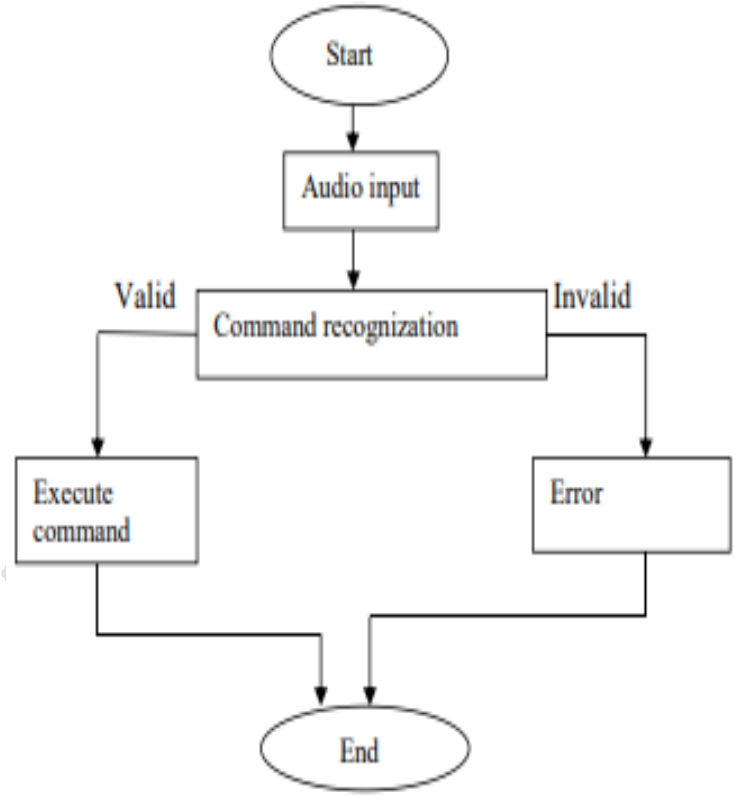

Fig 4.3: Voice Command Module

The Voice Command Module in the context of speech-to-text conversion is responsible for interpreting and processing voice commands from users, converting them into actionable instructions. This module typically combines speech recognition with natural language processing (NLP) to allow systems to respond to spoken commands in real time.

V. CONCLUSION

In conclusion, the project on converting speech to text and subsequently creating a PDF document has demonstrated the integration of advanced technologies to enhance accessibility, convenience, and productivity. By implementing Speech-to-Text (STT) conversion, the project has successfully bridged the gap between spoken language and written text, offering a valuable solution for users who prefer or require text- based information. The transformation of the transcribed text into a PDF document adds an extra layer of utility, making the information easily shareable, portable, and compatible across various platforms. In addition, these system uses the handicap persons for easily converting the speech into text also it is useful for the content creation and writing letter in less time. Also by using these system people can save their valuable time. Automating the transcription process saves time and effort, allowing users to focus on the content rather than the mechanics of typing. This can be particularly beneficial in business, education, and creative fields. Utilizing advanced speech recognition technology ensures that the transcriptions are accurate, reducing the need for

extensive editing. This reliability is crucial for maintaining the integrity of the recorded information. the "Convert Speech to Text and Create PDF" project successfully addresses the need for a comprehensive, user-friendly tool that enhances accessibility, efficiency, and document management

VI. REFERENCES

1) Rajat Saini, " Speech Recognition System (speech to text) (text to speech," IRJMETS, Volume:04/Issue:01/January-2022.

2) Sameer Gedam, Shubham Anand, Rakesh Badodekar Priya, Ashutosh Janbandhu," Informative Based Website Development Using Voice Commands," IJAER , Vol. No. 21, Issue No. IV,2021.

3) Babunditi, Dr.R.Praveen Sam, " Speech to Text Conversion using Deep Learning Neural Net Methods," TJCME, Vol.12/ No.05/2021

4) Nikhil Jain, Manya Goyal, Agravi Gupta, Vivek Kumar, " Speech to Text Conversion and Sentiment Analysis on Speaker Specific Data," IRJMETS, Volume:03/Issue:06/June-2021.

5) Shivangi Nagdewani, Ashika Jain, " A Review on Methods for Speech-To-Text And Textto-Speech Conversion,"IRJET, Volume: 07 Issue: 05/May 2020