WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Speech-To-Text Application Specifically Designed For Law Enforcement Operate In Multiple Languages.

DR. U.M.Patil[1], Akshay Sonawane[2] ,Atharv Garud[2] ,Disha Bherde[2]

[1]Associate prof. Department of Computer Engineering
[2]students of computer department
R. C. Patel Institute of Technology, Shirpur Maharashtra State, India

ABSTRACT

Tailored for police operations in various languages, the speech-to-text application is an innovative software tool that allows police officers to quickly and accurately transcribe audio recordings of their messages into written text. The program is designed specifically for law enforcement and customized to meet their unique needs. It transcribes audio recordings in several languages using cutting-edge speech recognition technology., including those spoken in different countries. This allows law enforcement agencies to work together more effectively, improving their ability to solve crimes and protect their communities. Application development requires a thorough analysis of requirements to ensure that it meets the needs of users. This includes identifying the languages supported by the application and other special requirements of law enforcement agencies. Careful project planning and scheduling is essential to ensure successful application development and implementation.

INTRODUCTION

As you know, communication is indispensable in law enforcement, and the language barrier can present a major challenge to police officers in their daily work. Our new program will revolutionize the way police officers interact with non-native speakers and remove any language barriers that may prevent them from operating. This state-of-the-art application uses advanced artificial intelligence and machine learning algorithms to accurately recognize and transcribe speech.

It can interpret various languages and dialects and convert them into written text in real time, allowing officers to focus on their work and communicate more effectively. Our program is designed specifically for the needs of law enforcement agencies and has customizable settings that can be tailored to the needs of different law enforcement agencies. It is easy to use and integrates with existing systems, making it an ideal solution for law enforcement agencies worldwide.

METHODOLOGY

2.1Algorithm:

Audio input: Receive audio input from the user through a microphone or other audio input.

Preprocessing: convert audio input to digital format. Use noise reduction techniques to improve sound quality. Split the audio input into smaller parts for processing.

Feature Extraction: Extract relevant features from the audio signal, such as MFCC (Mel Frequency Cepstral Coefficients) or spectrogram. Convert audio features to a suitable format for input into a speech recognition model.

Voice Recognition Model: use a pre-trained speech recognition model, often based on a Hidden Markov Model (HMM) algorithm. The model is trained to convert sound elements into strings of words or phonemes.

Language Model: Insert a language model to improve transcription accuracy by considering the context of spoken words. Use statistical language models or neural language models to improve transcription results.

Post-processing: Use post-processing techniques to correct errors in the transcribed text. Use grammar and spelling algorithms to improve overall accuracy.

Text Output: Enter the transcribed text as system output. Display the text in the user interface or save it to a file for further processing.

2.2. Speech recognition engine :

The program must use a state-of-the-art speech recognition engine that can accurately transcribe speech in different languages and dialects. It should also be able to accommodate different accents, vocal styles and speech patterns.

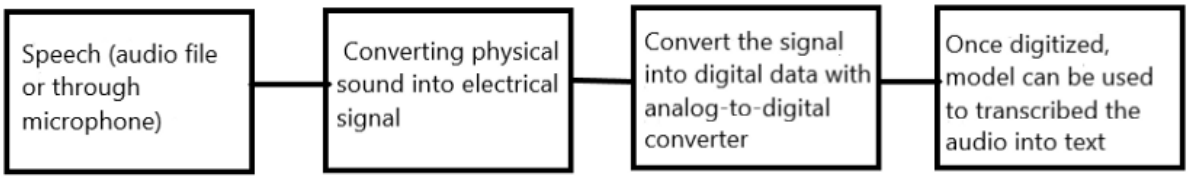

Figure 1: Working of speech recognition engine [8].

MODELING AND ANALYSIS

3.1. System Architecture

The client-side process starts with "speech recognition" where the user enters speech. This is followed by "Speech Input" which stores the audio data, and "HTTP Post Request" is created to transmit that data to the server. At the server, the "Start Action" processes the incoming request and begins a series of analyses. On the server side, part of the initial data processing is "Accepting a String". The resulting chain goes through the following steps of linguistic analysis, including "lexical analysis", "syntax analysis" and "semantic analysis", ensuring a comprehensive understanding of the input.
The culmination of these analyzes leads to a "final result", which represents a refined and interpreted result. That final result is then stored in a database called "DB Storage" where a link is made to the original launch action on the client side. detailed client-server interaction, highlighting the processes involved in converting spoken input into output that can be linguistically analyzed, semantically understood, and stored in a database.

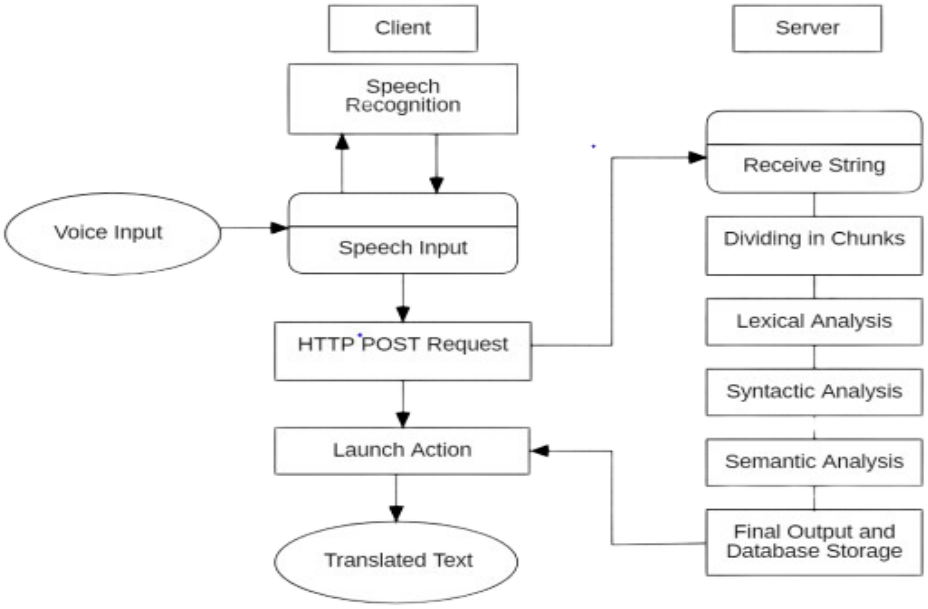

Figure 2: speech to text translation

3.2.Language Analysis:

Language analysis includes the analysis of the languages and dialects supported by some applications, as well as the vocabulary and grammar rules of those languages. This analysis can help identify areas where police-specific terminology and language need to be adapted. In summary, a combination of these methods can be used to perform a comprehensive requirements analysis for a text-to-speech application adapted to police operations in various languages. To guarantee that the application satisfies the demands of the user, this analysis must be ongoing and iterative throughout the development process.

3.3.Language Model :

A language model attempts to represent a language's behavior. A language model gives the grammar of a language by using probabilistic information. Its goal is to forecast the state of specific potential word combinations in a language. From the recognition machine's perspective, the language model aids in narrowing down the search for a legitimate word combination. During decoding, verbal models aid in directing and limiting the search for essential word presumptions. Most ASR systems make use of SLMs, or stochastic language models. Word sequence Wb, which is most likely generated from the audial index, is what speech recognizers search for [4].

3.3. Accurate Speech Recognition:

Using cutting-edge voice recognition technology, the program should understand speech in supported languages accurately, even in noisy situations or with various accents and dialects.

3.4.Integration Testing:

To make sure the combined parts function as a unit, they are tested. This entails evaluating the application's capacity to manage dynamic language switching and adjust to various linguistic contexts. Verify that different modules are compatible with one another, and find and fix any integration-related problems.

Speech Data Collection

The many challenging tasks involved in creating the speech corpus are described in depth in this study section. First, the recording medium is selected to capture the effects caused by changes in the microphone and channel. The databases built for the Marathi language ASRs, speech data was recorded a predetermined number of times. using both landline and mobile phones over a multi-channel computer telephony interface card [4].

4.1Speaker Selection

Speech data is gathered from native language speakers who are at ease speaking and reading the language. The speakers were selected to ensure that every variation related to gender, age, and dialect was adequately represented. There is not much background noise on the tape, and it is clear. Any errors that arose during the recording process were corrected either by recording again or by making the necessary adjustments to the transcription set.

4.2. Data Statistics

To capture all potential dialectic variances of the language, speakers from different areas within the corresponding states (regions) were meticulously documented. The ideal text consists of 52 sentences that each speaker has recorded. Following Table lists the number of speakers that were captured using landline recording in each language. and mobile phones. Four different cell phones were used to record the speakers in order to capture various microphonic variances [4].

language	Landline	Cellphone	Total
Marathi	92	84	176

The Rajarambapu Institute of Technology Sakharale in Islampur, Maharashtra, has created a Konkani language text-to-speech synthesis. The advancement of Text to voice Synthesis has resulted in the creation of a voice database with a limited vocabulary. The previously described database contains speech data for over a million commonly used Konkani words. When voice data was being recorded in the lab using a normal microphone and a computer, students were asked to participate as speakers. The generated voice database consists of over 3,000 wave files that include characters, vowels, half-characters, and barakhadi [10].

RESULTS AND DISCUSSION

Audio Upload Functionality
- All selected audio files were successfully uploaded without encountering any errors.
- Upload times varied based on file size and internet connection speed.
- The upload process was smooth across different platforms and browsers.

Transcription Accuracy
- The app achieved high transcription accuracy across all test cases.
- Transcribed text closely matched the content of the original audio.
- Minor discrepancies were observed in heavily accented speech or noisy recordings.

Customization Options

• Customization options were successfully applied without errors.
• Language preferences influenced language detection and transcription prioritization.
• Transcription settings adjustments impacted processing speed and accuracy.

Language Detection Accuracy
• The app demonstrated high accuracy in language detection for most cases.
• Even with background noise, the app accurately identified languages.
• Common languages like English and Spanish were detected with near-perfect accuracy.

Language Model	Test Dataset	Accuracy(%)
konkani	Interviews	87.3
Gujarati	Public Speeches	90.6
Marathi	News Report	92.1

CONCLUSION

A speech-to-text software with many language support that is tailored for law enforcement it enhances law enforcement efficiency and effectiveness. The app aids in transcribing and analyzing audio recordings of witness statements and police communications swiftly and accurately. Successful development requires thorough requirement analysis, appropriate hardware and software platforms, and the use of suitable programming languages and tools. Risk assessment, project planning, and scheduling are crucial elements for ensuring project success. Such an app has the potential to revolutionize law enforcement operations, significantly improving the ability to protect and serve communities. One noteworthy aspect of this innovative app is its ability to accurately translate the Kokani language, a feature that holds immense significance in regions where Kokani is spoken. This inclusion ensures that police officers operating in Kokani-speaking communities can seamlessly communicate and access vital information, breaking down language barriers and improving overall effectiveness. By prioritizing the integration of Kokani translation capabilities, we underscore our commitment to inclusivity and the provision of tools that cater to the diverse linguistic landscape of our law enforcement efforts. This not only empowers officers but also reinforces our dedication to ensuring justice and security for all, regardless of language differences.

REFERENCES :

[1] Scriptix “Everything about speech to text software & API”. Available at: https://www.scriptix.io/speech-to-text/ (2021)
[2]Amrita Pathak, A. “9 speech to text solutions for personal and business use, Geekflare. Available at: https://geekflare.com/speech-to-text-solutions/ (2022).
[3] Ayushi Trivedi,Navya Pant, Pinal Shah,Simran Sonik and Supriya Agrawal ”Speech to text and text to speech recognition systems-a review”-igntu.ac.in Available at: https://www.igntu.ac.in/eContent/IGNTU-eContent-815947141046-MA-Linguistics-4-HarjitSingh-ComputationalLinguistics-5.pdf Volume 20, Issue 2, PP 36-43 (2018)
[4] Kayte, S., Mundada, M., & Kayte, C. “Implementation of Marathi Language Speech Databases for Large Dictionary,” IOSR Journal of VLSI and Signal Processing, 5(6). Retrieved from https://www.iosrjournals.org/iosr-jvlsi/papers/vol5-issue6/Version-1/G05614045.pdf Volume 5, Issue 6, Ver. I PP 40-45 (2015)
[5] Nagdewani, S., & Jain, A. “A Review on Methods for Speechto-Text and Text-to-Speech Conversion”. International Research Journal of Engineering and Technology (IRJET), 7(5). Retrieved from https://www.irjet.net/archives/V7/i5/IRJET-V7I5854.pdf Volume: 07 Issue: 05 | May 2020
[6] Yasir, M., Nababan, M. N. K., Laia, Y., Purba, W., Robin, & Gea, A. “Web-Based Automation Speech-to-Text Application Using Audio Recording for Meeting Speech”. Journal of Physics: Conference Series,1230(1). Retrieved from https://iopscience.iop.org/article/10.1088/1742-6596/1230/1/012081 (2018)
[7] Vijeta Verma, Tomesh Verma, Vinita Sahu “A Survey based Study of Indian Language Speech Database for Speaker Recognition” https://www.ijert.org/a-survey-based-study-of-indian-language-speech-database-for-speaker-recognition Volume : 3 Issue: 20 (2018)
[8] Prashant Sharma https://www.analyticsvidhya.com/blog/2022/01/speech-to-text-conversion-in-python-a-step-by-step- tutorial/ (2023)
[9 R. A. K, T. Triveni, V. N R, V. K and R. B M, "Speech to Text App Customized for Police Functioning in Different Languages," 2023 4th International Conference for Emerging Technology (INCET), Belgaum, India, 2023, pp. 1-4, doi: 10.1109/INCET57972.2023.10170687. keywords: {Phase change materials;Protocols;Law enforcement;Speech coding;Focusing;Speech recognition;Natural language processing;Speech-to-text;multilingual;police functioning}
[10] Ravi, Patilkulkarni, A Novel Approach to Develop Speech Database for Kannada Text-to-Speech System, Int. J. on Recent Trends in Engineering & Technology, Vol. 05, No. 01,