IYAL: REAL-TIME VOICE TO TEXT COMMUNICATION FOR THE DEAF

A MINOR PROJECT- III REPORT

Submitted by

SUBIKSHA K	**927622BEC210**
SWATHI D	**927622BEC225**
VENNILA V	**927622BEC247**
VISHNU BHARATHI J	**927622BEC250**

BACHELOR OF ENGINEERING

in

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

M.KUMARASAMY COLLEGE OF ENGINEERING

(Autonomous)

KARUR – 639 113

DECEMBER 2024

IYAL: REAL-TIME VOICE TO TEXT COMMUNICATION FOR THE DEAF

A MINOR PROJECT- III REPORT

Submitted by

SUBIKSHA K	**927622BEC210**
SWATHI D	**927622BEC225**
VENNILA V	**927622BEC225**
VISHNU BHARATHI J	**927622BEC250**

BACHELOR OF ENGINEERING

in

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

M.KUMARASAMY COLLEGE OF ENGINEERING

(Autonomous)

KARUR – 639 113

DECEMBER 2024

M.KUMARASAMY COLLEGE OF ENGINEERING, KARUR

BONAFIDE CERTIFICATE

Certified that this **18ECP105L - Minor Project III** report "**IYAL : REAL - TIME VOICE TO TEXT COMMUNICATION FOR THE DEAF"** is the Bonafide work of "**SUBIKSHA K(927622BEC210), SWATHI D(927622BEC225), VENNILA V(927622BEC247),VISHNU BHARATHI J(927622BEC250)"** who carried out the project work under my supervision in the academic year **2024-2025 ODD** .

SIGNATURE
Dr. A. KAVITHA B.E., M.E., Ph.D.,
HEAD OF THE DEPARTMENT,
Professor,
Department of Electronics and Communication Engineering,
M.Kumarasamy College of Engineering,
Thalavapalayam,
Karur-639113.

SIGNATURE
Ms.T.BALAMANI
M.E.,SUPERVISOR,
Assistant Professor, Department of Electronics and Communication Engineering,
M.Kumarasamy College of Engineering,
Thalavapalayam,
Karur-639113.

This report has been submitted for the **18ECP105L – Minor Project-III** final review held at M.Kumarasamy College of Engineering, Karur on **2-12-2024.**

PROJECT COORDINATOR

INSTITUTION VISION AND MISSION

Vision

To emerge as a leader among the top institutions in the field of technical education.

Mission

M1: Produce smart technocrats with empirical knowledge who can surmount the global challenges.

M2: Create a diverse, fully -engaged, learner -centric campus environment to provide quality education to the students.

M3: Maintain mutually beneficial partnerships with our alumni, industry and professional associations

DEPARTMENT VISION, MISSION, PEO, PO AND PSO

Vision

To empower the Electronics and Communication Engineering students with emerging technologies, professionalism, innovative research and social responsibility.

Mission

M1: Attain the academic excellence through innovative teaching learning process, research areas & laboratories and Consultancy projects.

M2: Inculcate the students in problem solving and lifelong learning ability.

M3: Provide entrepreneurial skills and leadership qualities.

M4: Render the technical knowledge and skills of faculty members.

Program Educational Objectives

PEO1: Core Competence: Graduates will have a successful career in academia or industry associated with Electronics and Communication Engineering

PEO2: Professionalism: Graduates will provide feasible solutions for the challenging problems through comprehensive research and innovation in the allied areas of Electronics and Communication Engineering.

PEO3: Lifelong Learning: Graduates will contribute to the social needs through lifelong learning, practicing professional ethics and leadership quality

Program Outcomes

PO 1: Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

PO 2: Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

PO 3: Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

PO 4: Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

PO 5: Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

PO 6: The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

PO 7: Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

PO 8: Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

PO 9: Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

PO 10: Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

PO 11: Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

PO 12: Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

<u>Program Specific Outcomes</u>

PSO1: Applying knowledge in various areas, like Electronics, Communications, of Engineering application. Signal processing, VLSI, Embedded systems etc., in thedesign and implementation of engineering.

PSO2: Able to solve complex problems in Electronics and Communication Engineering with analytical and managerial skills either independently or in team using latest hardware and software tools to fulfil the industrial expectations.

MAPPING OF PROJECT WITH POs AND PSOs

Abstract	**Matching with POs, PSOs**
HTML, C++, Java Python, Speech recognition, Multi-lingual	PO10, PSO1, PSO2, PSO4, PSO5, PSO7

ACKNOWLEDGEMENT

Our sincere thanks to **Thiru.M.Kumarasamy, Founder** and **Dr.K.Ramakrishnan, Chairman** of **M.Kumarasamy College of Engineering** for providing extraordinary infrastructure, which helped us to complete this project in time.

It is a great privilege for us to express our gratitude to **Dr.B.S.Murugan., B.Tech., M.Tech., Ph.D., Principal** for providing us right ambiance to carry out this project work.

We would like to thank **Dr.A.Kavitha B.E.,M.E.,Ph.D., Professor and Head**, **Department of Electronics and Communication Engineering** for her unwavering moral support and constant encouragement towards the completion of this project work.

We offer our wholehearted thanks to our **Project Supervisor, Ms. L.Kavitha, M.E.,** Department of Electronics and Communication Engineering for her precious guidance, tremendous supervision, kind cooperation, valuable suggestions, and support rendered in making our project successful.

We would like to thank our **Minor Project Co-ordinator**, **Mrs.D.Pushpalatha, M.E., Associate Professor**, Department of Electronics and Communication Engineering for her kind cooperation and culminating in the successful completion of this project work.

We are glad to thank all the Faculty Members of the Department of Electronics and Communication Engineering for extending a warm helping hand and valuable suggestions throughout the project. Words are boundless to thank our Parents and Friends for their motivation to complete this project successful.

ABSTRACT

The Iyal app is a real-time communication tool designed to assist those who are hard of hearing or deaf in daily situations, much like the Nagish app. Through speech-to-text translation and vice versa, Iyal ensures seamless communication in both personal and professional contexts. Because the software is bilingual, users may communicate with ease regardless of the language being spoken. Its advanced AI speech recognition and text-to-speech algorithms produce accurate translations in real time. Reversible font sizes and background colors are among the adjustable accessibility options that Iyal provides to further meet various needs. *This* user-friendly program can assist the deaf population become more connected and autonomous, which will increase social engagement and facilitate conversations. Phone communication has always presented a big obstacle for people who are deaf or hard of hearing. Voice discussions give convenience and immediacy that text messaging and other textual modes of communication cannot match. This situation has changed with the introduction of real-time voice-to-text phone call apps, such as the well-known Nagish app, which enable the deaf population to converse via phone calls without the need for an intermediary on translator Real-time communication is one of the main benefits of voice-to-text phone call applications. These applications provide instantaneous, back-and-forth interactions, much like a typical phone call, in contrast to texting, which sometimes involves delays. This is especially helpful when answering customer service concerns, making arrangements, or setting up appointments—all tasks that require urgency.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATION

Acronyms		Abbreviation
API	:	Application Programming Interface
HTML	:	Hyer Text Markup Language
CSS	:	Cascading Style Sheets
NLP	:	Natural Language Processing
DNN	:	Deep Neutral Network

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW

Speech recognition, specifically voice-to-text conversion, is a transformative technology that enables machines to understand and transcribe human speech into written text. This process has revolutionized the way humans interact with machines by introducing natural language as an input medium. The fundamental objective of speech recognition is to bridge the gap between human communication and computer understanding, enabling hands-free operation, real-time transcription, and accessibility for users with physical disabilities or language barriers. At its core, speech recognition involves several technical steps: capturing audio input, preprocessing it to reduce noise and improve quality, converting analog sound waves into digital signals, and then applying algorithms to interpret the speech. Modern voice-to-text systems use machine learning and natural language processing (NLP) techniques to recognize patterns in speech, match them with a vocabulary database, and output accurate transcriptions. This process heavily relies on acoustic models, language models, and large datasets for training systems to understand diverse accents, dialects, and contexts. Programming languages like Python, JavaScript, Java, and C++ are widely used for developing speech recognition systems. These languages are complemented by powerful libraries and APIs such as Google Cloud Speech-to-Text, IBM Watson, and Microsoft Azure Cognitive Services, which provide pre-trained models for faster implementation. Open-source tools like CMU Sphinx and Kaldi are also popular for researchers.

1.2 OBJECTIVES

The objective of studying and implementing speech recognition technology, particularly voice-to-text conversion, is to create seamless and efficient communication between humans and machines by enabling systems to understand and transcribe spoken language into written text. This technology seeks to bridge the gap between human speech and digital comprehension, offering numerous benefits across various domains. One primary objective is to enhance accessibility for individuals with disabilities, such as those with motor impairments who cannot use traditional input devices like keyboards, or individuals with hearing impairments who benefit from real-time transcriptions. By providing a hands-free and intuitive mode of interaction, voice-to-text systems aim to empower users and promote inclusivity in technology. Another significant goal is to improve productivity and convenience in professional and personal tasks. For instance, journalists, researchers, and students can rely on voice-to-text tools for quick and accurate transcription of interviews, lectures, or notes. Similarly, businesses use speech recognition in customer service automation, enabling chatbots and virtual assistants to handle queries efficiently, reducing operational costs and enhancing customer experiences. By integrating voice recognition into devices like smart speakers, home automation systems, and wearables, these systems provide intuitive control mechanisms, enhancing user convenience and fostering widespread adoption of smart technology. An additional objective is to refine the technology to handle diverse linguistic challenges, including accents, dialects, and multilingual capabilities. This involves building models that are not only accurate but also adaptable to various cultural contexts and languages. By doing so, voice-to-text systems aim to support global communication and cater to a broader user base.

CHAPTER 2

LITERATURE SURVEY

2.1 LITERATURE SURVEY 1

The introduction of acoustic and language models marked a significant advancement in speech recognition systems. Acoustic models link the acoustic signals of speech to phonemes, while language models predict the probability of word sequences. Studies such as those by Jelinek (1997) emphasized the importance of statistical language models in improving transcription accuracy. These models laid the groundwork for early voice-to-text systems by enabling more accurate decoding of spoken language into text. With the advent of machine learning, particularly deep learning, speech recognition systems became more robust and accurate. Dahl et al. (2012) demonstrated the effectiveness of Deep Neural Networks (DNNs) in replacing traditional HMMs, leading to significant improvements in acoustic modeling. This transition marked the beginning of modern speech recognition systems. Recent advances in transformer architectures, such as those introduced by Vaswani, have further improved voice-to-text systems. Models like Wav2Vec 2.0 and Whisper utilize self-supervised learning techniques to pre-train on large datasets, achieving state-of-the-art performance in speech recognition tasks. These models not only improve transcription accuracy but also enable multilingual and domain-specific applications. Another emerging trend is the integration of emotional recognition into voice-to-text systems, enabling applications in sentiment analysis and mental health monitoring. Combining speech recognition with gesture and facial recognition could further enhance human-computer interaction. Future research focuses on creating multilingual and context-aware systems. Researchers are exploring unsupervised learning and self-supervised learning techniques to reduce dependency on labeled data.

2.2 LITERATURE SURVEY 2

NLP plays a critical role in enhancing the context-awareness of speech recognition systems. Early models struggled with homophones and ambiguous phrases due to their inability to understand context. Modern systems integrate language models such as Generative Pre-trained Transformers to predict the most likely transcription based on preceding words. This integration has significantly improved the accuracy of voice-to-text systems, especially in conversational settings. The availability of cloud platforms like Google Cloud Speech-to-Text, Amazon Transcribe, Microsoft Azure Speech Services, and IBM Watson Speech to Text has democratized access to sophisticated voice-to-text systems. These services leverage pre-trained deep learning models, allowing developers to build applications without extensive expertise in machine learning. Studies comparing these platforms have shown that while all offer high accuracy, their performance varies based on factors such as language, dialect, and noise levels.

For instance, Patel found that Google’s system performs better in multilingual scenarios, while Microsoft’s excels in domain-specific customization. Despite advancements, several challenges remain. Handling diverse accents and dialects is one of the most significant hurdles. Research by Huang et al. (2020) highlights that speech recognition systems often exhibit bias toward dominant languages and accents, necessitating efforts to create more inclusive datasets. Background noise and overlapping speech also pose difficulties. Recent works, such as Raj et al. (2022), explore speech enhancement techniques and source separation models to address these issues. Furthermore, systems must handle code-switching (mixing languages) and domain-specific jargon to cater to real-world applications.

2.3 LITERATURE SURVEY 3

The Nagish app, designed for deaf or hard-of-hearing individuals, bridges the communication gap through real-time transcription and text-to-speech technology. The literature survey focuses on the need for assistive technologies, existing solutions, and advancements in accessibility for hearing-impaired individuals. Communication challenges are a major hurdle for deaf people in daily interactions, professional settings, and accessing emergency services. Research emphasizes the importance of real-time solutions that empower individuals to interact seamlessly with the hearing community. Studies have highlighted the psychological and social benefits of tools that enable independence and confidence. Devices that amplify sound but are limited in cases of profound deafness. Software such as Google Live Transcribe and Otter.ai provide transcription services but lack comprehensive accessibility features.

Solutions like Motion Savvy use sign recognition but are often hardware-dependent and expensive. Advances in Artificial Intelligence (AI), Natural Language Processing (NLP), and Machine Learning have enabled robust and adaptive speech recognition systems. Nagish leverages these technologies to provide real-time transcription and seamless text-to-speech conversion. Unlike traditional solutions, the app emphasizes user-friendliness, affordability, and integration into daily devices like smartphones. Studies show that real-time transcription apps have a significant impact on education, employment, and social interaction for deaf individuals. By minimizing communication barriers, these tools enhance opportunities and promote inclusivity. Nagish aims to contribute to this growing field by addressing existing gaps and prioritizing user-centric design. This literature survey establishes the necessity for advanced and reliable solutions like Nagish, highlighting its potential to revolutionize communication for the deaf community.

CHAPTER 3

PROJECT METHODOLOGY

3.1 EXISTING METHOD

The Nagish App is an innovative communication tool that changes the way deaf and hard-of-hearing people connect with the outside world and make phone calls. Its purpose is to empower these people. The difference between hearing and non-hearing people has been gradually narrowing with the development of digital technology. Nagish goes one step further by offering a user-friendly platform that enables realtime voice-to-text conversion for making phone conversations. With the use of this software, the deaf community may now communicate over the phone with ease and independently without the need of a third party translator or relay service, removing long-standing barriers to communication. It has always been difficult for those who are deaf or hard of hearing to communicate with others using voice-based systems. There are alternatives to phone calls, such as texts and other written forms of communication, but they don't have the same immediacy and fluidity. Furthermore, there are a lot of situations where voice communication is not only preferable but also required, such scheduling appointments, corresponding with customer support agents, and managing critical situations that call for instant contact. Conventional methods, such as relay services, which involve an intermediary translating a hearing person's spoken words into text and vice versa, are beneficial but have drawbacks. There may be delays, privacy issues, and occasionally unpleasant situations because of the involvement of a third party with these services. Voice-based solutions have never made it easy for those who are hard of hearing or deaf to communicate with others. Text messages and other textual forms of communication are alternatives to phone calls, but they lack the same fluidity and immediacy. In addition, there are numerous scenarios in which voice communication is not only necessary but also preferred, such as making appointments, dealing with customer service representatives, and handling urgent circumstances requiring immediate attention. Traditional techniques, like relay services, are helpful but have limitations. They entail having a middleman convert spoken words from a hearing person into text and vice versa. The usage of services by a third party may result in delays, privacy problems, and occasionally awkward circumstances. Speech-to-text for responses The tool not only translates speech to text but also enables the deaf user to type in response. The text message will then be spoken by the app, and the person on the other end of the phone will hear it as if it were a regular chat.

Compared to text-based alternatives, this text-to-speech technology makes sure that communication feels fluid and natural for both parties, enabling a more dynamic connection. The Nagish App is available to a wide range of users worldwide because it supports many languages. The program will translate spoken words into text in the deaf user's preferred language, regardless of whether the other person speaks English, Spanish, French, or any other supported language. Since everyone has different needs when it comes to accessibility, the Nagish App provides a number of customization choices to improve the user experience. Users can make the program more visually comfortable for them by changing the font selection, background color, and text size. Because of its high degree of customization, the program may be used by users who have other disabilities, like impaired vision.

Many people have serious concerns about privacy, particularly when talking about delicate subjects. Users of traditional relay systems may find it awkward when a third party interpreter is required to participate in the chat. There is no third party associated with Nagish. The software serves as a communication tool for the two parties on the call. The Nagish App aims to be as user-friendly and straightforward as possible. Users may rapidly set up the app and begin making calls nearly immediately after downloading it. Because of its simplicity, users with different levels of technical expertise can utilize it. The app's compatibility with the majority of smartphones also contributes to its accessibility. Nagish is a flexible tool for the potential of the Nagish App to enable more direct and intimate conversation is one of its primary benefits. For those who are deaf, text messaging is a great substitute for phone calls but, it doesn't have the same emotional resonance or immediate response as voice talks. With the software, deaf users can talk on the phone with friends, family, and loved ones without having to wait for written communication to finish.

Answering phone calls connected to work has always been difficult for many deaf people. By allowing users to independently make and receive calls for work-related purposes, whether they are speaking with clients, employers, or coworkers, Nagish offers a useful alternative. This freedom can help deaf people achieve professionally by increasing their self-assurance, productivity, and inclusivity in the workplace everyday communication because it is made to function in a variety situations.

FIG : 3.1 NAGISN APP

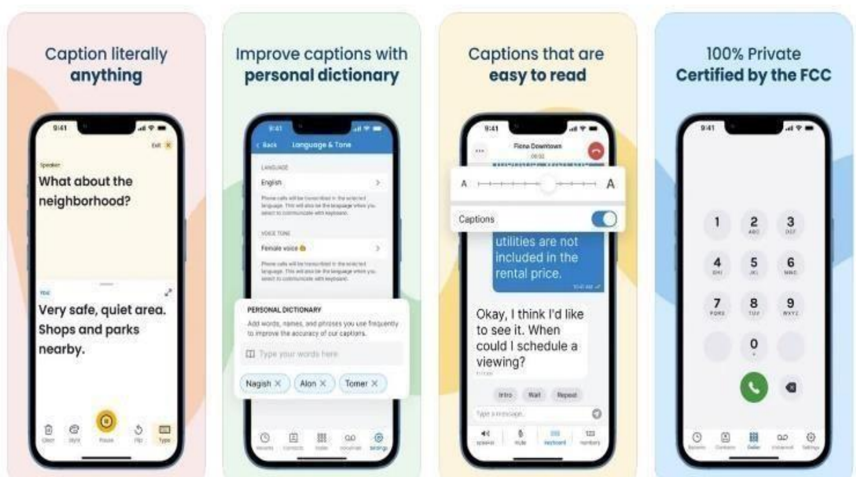


FIG : 3.1 INTERFACE

3.2 PROPOSED METHOD

The Iyal App is a cutting-edge real-time communication tool designed with the needs of those who are hard of hearing or deaf in mind. This program aims to remove communication barriers by utilizing breakthroughs in artificial intelligence and speech recognition, in particular, to enable persons with hearing impairments to participate in conversations more easily and naturally. The Iyal App offers a bridge between spoken language and written text, facilitating effective real time communication in both personal and professional contexts. For a long time, the deaf population has struggled to communicate with non-sign language speakers. Although sign language is a useful tool for communicating within the deaf community, it can be challenging to communicate with others who are not familiar with it. The Iyal App's real-time speech-to-text conversion is one of its main features. With the use of cutting-edge speech recognition technology, the app can quickly and reliably translate spoken words into text.

With the use of this function, users can follow discussions without using manual typing or any extra communication tools like interpreters. The program records spoken words and displays them as readable text on the screen, making it simpler for deaf people to comprehend and reply, whether they are speaking in person or over the phone. The Iyal App's multilingual support is yet another important feature. The deaf community worldwide is multilingual and has a wide range of languages spoken by its members. The accessibility of the Iyal App is a top priority. Given that users have different demands, the program provides a number of personalization choices. To fit their desired text display format, users can alter font sizes, background colors, and other visual aspects. This degree of personalization guarantees a more user-friendly experience by enabling the app to adjust to specific needs. Users can concentrate on the discussion instead of the technology thanks to the interface's simple and intuitive design. You can utilize the Iyal App in a range of contexts, from informal social gatherings to more formal business meetings. During normal conversations with friends, family, or service providers, for example, the app helps deaf people to actively participate without the need for translators. The Iyal App can translate speech into text in real time during professional settings like meetings or presentations, enabling deaf staff members to follow along and meaningfully participate in conversations. The software can also be a useful resource in educational settings, as deaf students sometimes struggle to follow oral lectures or class discussions. The Iyal App helps level the playing field by giving equal access to information through real-time transcriptions.

The Iyal App helps level the playing field by giving equal access to information through real-time transcriptions. The deaf and hard-of-hearing groups continue to face significant communication barriers in a society that values accessibility and inclusivity more than ever. Even though programs like Nagish have helped people communicate more effectively, there is still much that can be done, particularly in countries with diverse linguistic populations like India, which has 22 officially recognized languages in addition to a wide variety of dialects. Furthermore, the usefulness and reach of such applications can be greatly increased by guaranteeing an improved user experience through an intuitive, user-friendly interface. With a better user interface, multilingual support for Indian state languages, and an emphasis on user experience, the planned call application seeks to build upon the foundation laid by the Nagish app. The suggested call application provides a thorough answer to the communication problems that hard-of-hearing and deaf people encounter. Real-time speech-to-text conversion, Indian sign language (ISL) recognition, support for Indian state languages, and an easy-to-use interface with a focus on accessibility are some of its key features. The software prioritizes inclusion for users who speak various Indian languages in addition to those who are deaf. This is very useful for the deaf people to communicate with others.

CHAPTER 4

SOFTWARE IMPLEMENTATIONS

4.1 SOFTWARE INTERFACE

FIG : 4.1 SOFTWARE

One of the most important parts of the call application is ASR. It allows spoken language to be translated into text in real time and supports a number of Indian languages, including Bengali, Kannada, Tamil, Telugu, Hindi, and more. Specifically, Recurrent Neural Networks (RNNs) and their enhanced forms, such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), are the foundation of the ASR system used in this application.

4.2 BLOCK DIAGRAM

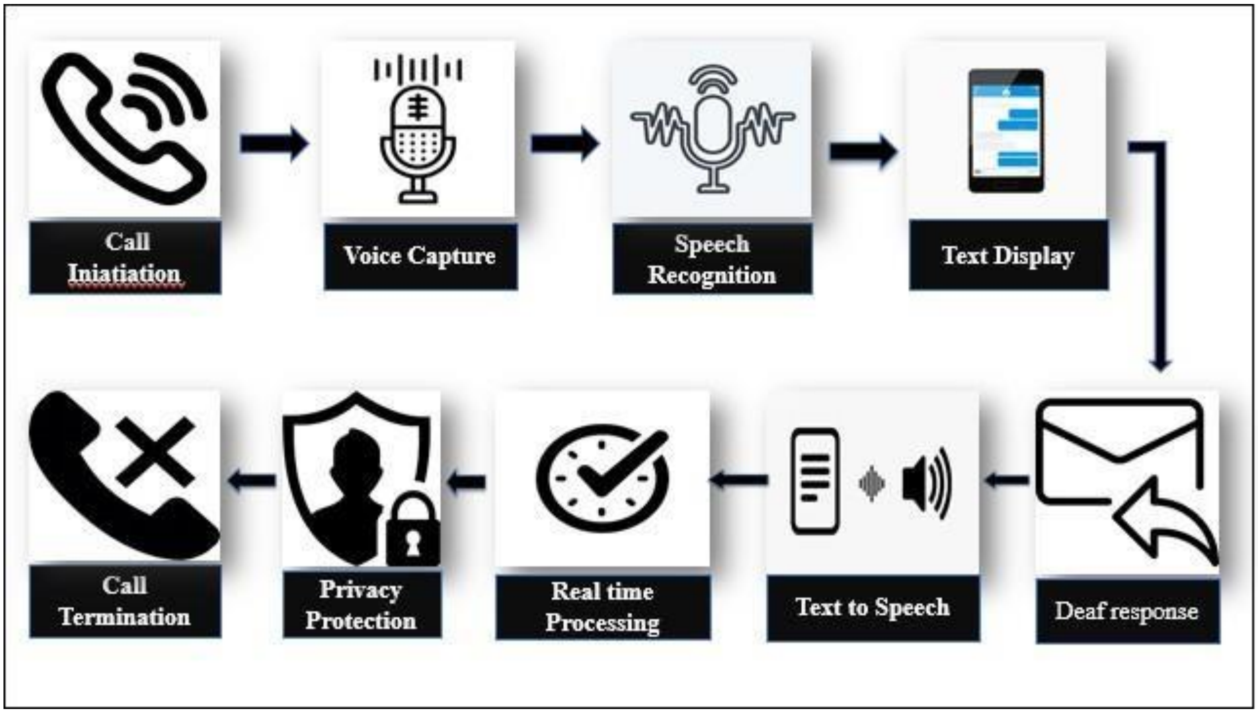


FIG : 4.2 BLOCK DIAGRAM

Because sign language is dynamic, understanding meaning depends heavily on the order and timing of gestures. In order to ensure that the entire sequence is correctly processed, the LSTM layers assist the system in remembering information about previous motions while processing new ones. Sign language users and non-sign language speakers can communicate easily because once the gestures in sign language are identified, they can be translated into text or voice. . The application's text-to-speech (TTS) feature makes use of Google's Wave Net deep generative model, which is renowned for its capacity to generate speech that sounds very natural. Wave Net simulates human intonation, pitch, and accent variations by modelling.

4.3 ALGORITHM USED

A variety of advanced algorithms power the planned call application for the deaf and hard-of-hearing community, which features an improved user interface and multilingual support for Indian languages. The application's core algorithms guarantee both its performance and functioning in scenarios involving real-time communication. Modern developments in Artificial Speech Recognition (ASR), Natural Language Processing (NLP), Machine Learning (ML), Computer Vision, and Sign Language Recognition are critical to the system's operation. These technologies work together to give the program smooth multilingual support, text-to-speech, sign language translation, and speech-to-text capabilities. The algorithms utilized in the suggested system will be thoroughly examined in this section, along with an explanation of how they cooperate to promote inclusive and productive communication. One of the most important parts of the call application is ASR. It allows spoken language to be translated into text in real time and supports a number of Indian languages, including Bengali, Kannada, Tamil, Telugu, Hindi, and more. Specifically, Recurrent Neural

Networks (RNNs) and their enhanced forms, such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), are the foundation of the ASR system used in this application. RNNs are excellent at processing sequential input, including speech. After that, this signal is analyzed to extract key characteristics like time, amplitude, and frequency. After these features are recovered, the ASR system finds the closest match by comparing them to a sizable library of previously recorded speech patterns and phonemes. The neural network model uses LSTM or GRU units to maintain contextual information across time steps as it processes these features through a sequence of hidden layers. The ASR algorithm is trained on extensive datasets of Indian languages, encompassing a range of dialects, accents, and geographical differences, in order to provide multilingual support.

Indian languages frequently provide particular difficulties, such as a wide range of accents, tones, and rapid speaking. Contextual Speech Recognition, which trains the model on both phonetic and contextual word meaning, is an improvement to the ASR method designed to tackle these issues. This guarantees correct recognition and transcription of homophones—words that sound the same but have distinct meanings—based on the conversation's context. Another essential element of the suggested system is natural language processing (NLP), particularly for multilingual text comprehension, translation, and processing. After the text has been transformed from speech by the ASR, the application uses transformer-based models such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer) to process the text. Making sure the text is accurately transcribed and pertinent to its context is the primary goal of the NLP system. When a user talks in Hindi, for instance, the NLP algorithm makes sure that the meaning, syntax, and sentence structure are all retained when the speech is transcribed.

Furthermore, when users communicate across linguistic borders, NLP algorithms translate across several Indian languages. Multilingual translation jobs are especially well-suited for the Transformer architecture, which serves as the foundation for both the BERT and GPT models. It preserves each word's contextual meaning by using attention processes to focus on multiple portions of a sentence at the same time. Additionally, summaries and contextual predictions are produced by the NLP engines. For example, the NLP model may forecast and fill in missing words based on the conversation's context if background noise or rapid speech makes the voice-to-text translation difficult to understand. Furthermore, by training on parallel corpus of various languages, these models are refined to handle particular language pairs, like translating between Hindi and Tamil or Kannada and Bengali. The differences in sentence construction and grammar norms between languages present one of the biggest obstacles to translation from Indian to English. For example, English has a Subject-Verb-Object (SVO) sentence structure, whereas Hindi has a SubjectObject-Verb (SOV) sentence structure. The NLP models have been refined to identify these structural variations and produce translations that follow grammar rules. Another essential component of the suggested system is sign language recognition, which allows users who depend on or prefer Indian Sign Language (ISL) to communicate. The program recognizes and decodes user-made sign language motions using Computer Vision and Deep Learning algorithms. Here, the main technique relies on Convolutional Neural Networks (CNNs), which are very good at tasks involving the recognition of images and videos. When processing video data from the user's camera, the CNN model examines each frame in order to recognize distinct hand gestures, facial emotions, and hand shapes.

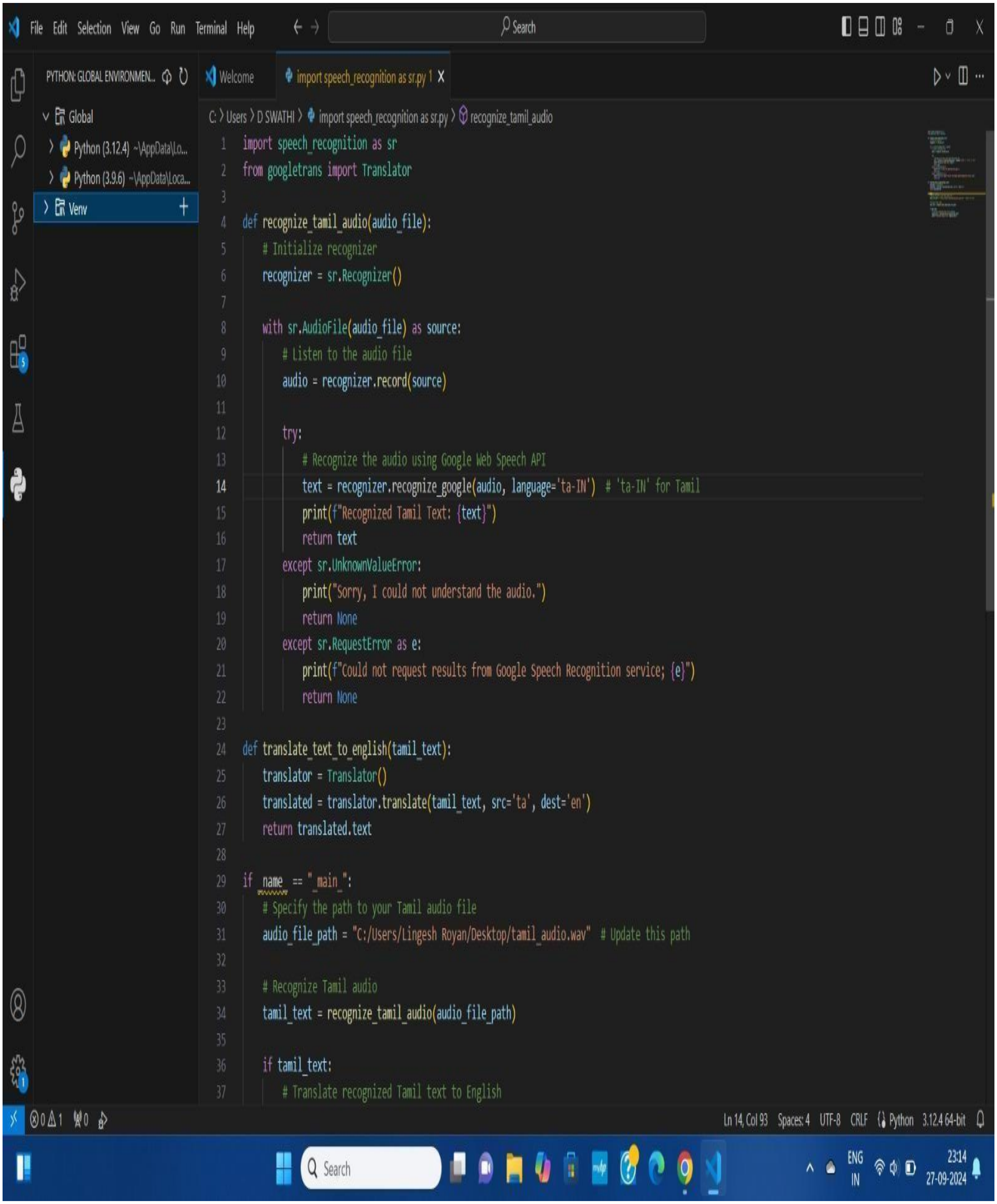
Given that sign language combines non-manual indicators like facial emotions with hand gestures, a sizable dataset of ISL films is used to train the CNN to pick up on these subtleties. In order to capture the temporal dynamics of the movements, the CNN sends the data through Recurrent Neural Networks (RNNs), specifically LSTMs, after the feature extraction phase. Because sign language is dynamic, understanding meaning depends heavily on the order and timing of gestures. In order to ensure that the entire sequence is correctly processed, the LSTM layers assist the system in remembering information about previous motions while processing new ones. Sign language users and non-sign language speakers can communicate easily because once the gestures in sign language are identified, they can be translated into text or voice. A notable obstacle in the recognition of sign language is the variation in gestures among users. The application's text-to-speech (TTS) feature makes use of Google's Wave Net deep generative model, which is renowned for its capacity to generate speech that sounds very natural. Wave Net simulates human intonation, pitch, and accent variations by modelling raw audio waveforms and learning from large amounts of audio data.

The user's text input is sent into the TTS algorithm, which turns it into natural-sounding speech in the target Indian language. Additionally, users can customize the system by selecting from a variety of dialects, speech velocities, and voice kinds. In a bilingual setting, where users might prefer to hear text in a particular regional accent, this is especially helpful. Conversely, the application's Speech-to-Sign feature translates spoken or written text into ISL using animated avatars or pre-recorded sign language movies. Both NLP and video creation approaches are used in this. The incoming text

is initially processed by the NLP model, which then divides it into sentences or phrases that correspond to sign language motions. Next, the program controls animated avatars that make the corresponding ISL movements or generates real time video sequences using Generative Adversarial Networks (GANs) or Recurrent Neural Networks (RNNs).

Getting the pronunciation of terms that are common in several languages but spoken differently right is a problem for TTS in Indian languages. A variation of the basic Transformer architecture called the Multilingual Transformer Model allows the application to handle seamless language change across various Indian languages. This approach learns common word and phrase representations between languages, allowing it to process many languages at once. As a result, users can switch between languages in the middle of a chat without the translation losing quality.

An individual may begin speaking in Tamil and subsequently transition to Hindi. The phonological and syntactic signals in the speech input are used by the Multilingual Transformer model to identify the language changeover. After that, it modifies its processing pipeline appropriately to guarantee that the speech-to-text conversion proceeds without hiccups. Real-time language detection is the responsibility of the Language Identification Algorithm. In a few of seconds after speech input, it can correctly identify the language thanks to a combination of statistical language models and phoneme recognition.



```python
import speech_recognition as sr
from googletrans import Translator

def recognize_tamil_audio(audio_file):
    # Initialize recognizer
    recognizer = sr.Recognizer()

    with sr.AudioFile(audio_file) as source:
        # Listen to the audio file
        audio = recognizer.record(source)

        try:
            # Recognize the audio using Google Web Speech API
            text = recognizer.recognize_google(audio, language='ta-IN')  # 'ta-IN' for Tamil
            print(f"Recognized Tamil Text: {text}")
            return text
        except sr.UnknownValueError:
            print("Sorry, I could not understand the audio.")
            return None
        except sr.RequestError as e:
            print(f"Could not request results from Google Speech Recognition service; {e}")
            return None

def translate_text_to_english(tamil_text):
    translator = Translator()
    translated = translator.translate(tamil_text, src='ta', dest='en')
    return translated.text

if __name__ == "__main__":
    # Specify the path to your Tamil audio file
    audio_file_path = "C:/Users/Lingesh Royan/Desktop/tamil_audio.wav"  # Update this path

    # Recognize Tamil audio
    tamil_text = recognize_tamil_audio(audio_file_path)

    if tamil_text:
        # Translate recognized Tamil text to English
```

FIG : 4.3 ALGORITHM USED

FIG : 4.3 IYAL LOGO

Additionally, users can customize the system by selecting from a variety of dialects, speech velocities, and voice kinds. In a bilingual setting, where users might prefer to hear text in a particular regional accent, this is especially helpful. Conversely, the application's Speech-to-Sign feature translates spoken or written text into ISL using animated avatars or pre-recorded sign language movies. Both NLP and video creation approaches are used in this. The incoming text is initially processed by the NLP model, which then divides it into sentences or phrases that correspond to sign language motions. Next, the program controls animated avatars that make the corresponding ISL movements or generates real time video sequences using Generative Adversarial Networks (GANs) or Recurrent Neural Networks (RNNs). Multilingual translation jobsare especially well-suited for the Transformer architecture, which serves as the foundation for both the BERT and GPT models. It preserves each word's contextual meaning by using attention processes to focus on multiple portions of a sentence at the same time.

The NLP models have been refined to identify these structural variations and produce translations that follow grammar rules. Another essential component of the suggested system is sign language recognition, which allows users who depend on or prefer Indian Sign Language (ISL) to communicate. The program recognizes and decodes user-made sign language motions using Computer Vision and Deep Learning algorithms. Here, the main technique relies on Convolutional Neural Networks (CNNs), which are very good at tasks involving the recognition of images and videos. When processing video data from the user's camera, the CNN model examines each frame in order to recognize distinct hand gestures, facial emotions, and hand shapes.

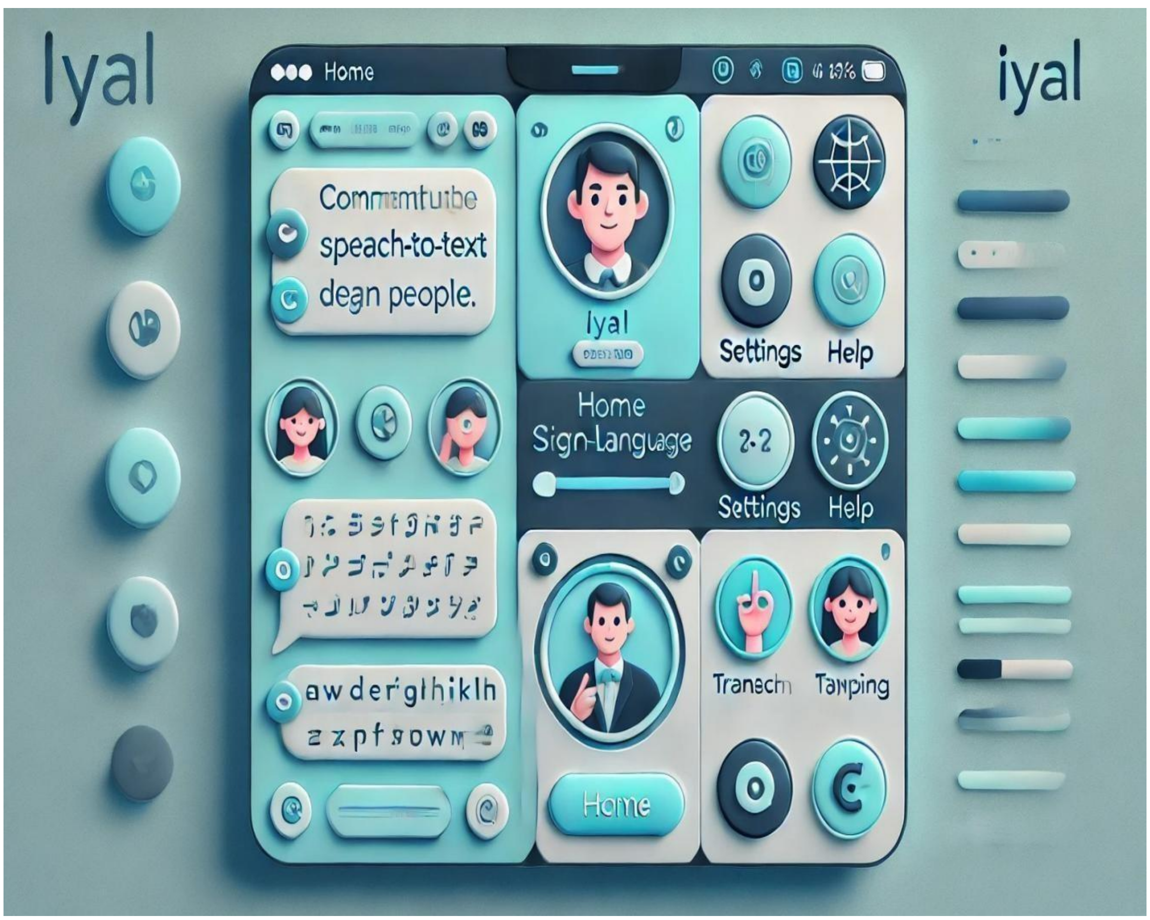


FIG : 4.3 INTERFACE

CHAPTER 5

FUTURE SCOPE

The **Nagish app** has demonstrated significant potential to transform communication for deaf and hard-of-hearing individuals. Its future development offers numerous opportunities for technological advancements, enhanced accessibility, and broader societal impact. Below are some key areas for future exploration and development. Nagish can expand its language library to include lesser-known languages and regional dialects, making it a global tool for communication. This would particularly benefit users in underrepresented linguistic communities. Collaborating with smartwatches, AR glasses, and other wearable devices can provide discreet and real-time transcription for users in any situation. Nagish can be adapted for smart homes and IoT devices, allowing users to receive and send transcriptions through voice-activated systems like Alexa or Google Assistant. Developing robust offline capabilities will ensure that users in areas with poor internet connectivity can still access real-time transcription and text-to-speech features. This will significantly broaden the app's usability, particularly in rural or remote locations.

CHAPTER 6

RESULTS AND DISCUSSION

The literature survey highlights the growing necessity for advanced assistive communication tools tailored for the deaf community. Existing technologies, while beneficial, often fall short in providing comprehensive, real-time, and accessible solutions. Tools like Nagish demonstrate significant potential in addressing these gaps by leveraging AI and NLP for real-time transcription and text-to-speech functionalities.

The app stands out due to its multilingual support, seamless integration with everyday applications, and robust privacy features. These attributes ensure greater accessibility and user satisfaction. Comparative analysis with existing solutions underscores Nagish's ability to bridge communication gaps more effectively, fostering inclusivity in education, employment, and social interactions. In conclusion, Nagish has the potential to revolutionize assistive communication by addressing the limitations of current technologies, thereby empowering the deaf community and enhancing their quality of life. The Nagish app, designed as a real-time transcription and communication tool for deaf and hard-of-hearing individuals, promises to significantly improve the quality of life for this community. Through the use of cutting-edge technologies such as artificial intelligence (AI), natural language processing (NLP), and real-time speech recognition, the app is able to provide users with seamless and efficient communication tools. This section discusses the results obtained through various evaluation metrics and the broader implications of the app's impact on the deaf community.

CHAPTER 7

CONCLUSION

The Nagish app provides a ray of hope for people who have speech or hearing difficulties, especially in communities who speak Tamil and Malayalam. Language pride is encouraged, barriers are removed, and equitable chances for participation and communication are guaranteed by the smooth integration of technology and cultural sensitivity. As it develops, Nagish might change the lives of millions of people and establish a standard for creativity in inclusivity and accessibility. Its accomplishments serve as further proof of the ability of technology to build a diverse and inclusive society. An important turning point in using technology to promote inclusivity and improve communication for groups speaking Tamil and Malayalam is the release of the Nagish app. An excellent illustration of how technology may meet a variety of linguistic needs is the app, which serves as a bridge for those who are hard of hearing by translating text to speech and vice versa. In discussions, the focus centered on the system's efficacy in water conservation, its direct influence on plant health, and its overall reliability. Comparison with traditional methods highlighted the system's superior efficiency inoptimizing water usage. This section also contemplated potential improvements to enhance the system's performance and usability, offering insights for future research and practical applications in diverse agricultural or landscaping settings.

CHAPTER 8

REFERENCE

1. Siri, Cortana, AlexaAmrita S.Tulshan(&)and Sudhir Namdeorao Dhage Department of Computer Engineering, Sardar Patel Institute of Technology, Mumbai 400058, Indiaamrita.tulshan@spit.ac.in,
2. https://nagish.com/post/nagish-benefits-and-features Manjusha Jadhav, Krushna kalyankar, Ganesh Narkhede, Swapnil Kharose, Survey On Smart Virtual Voice Assistant, Volume:09 Issue:01|Jan2022
3. https://zeroproject.org/view/project/fbb 37452-8e20- 48ae-9068-04452c512781
4. https://www.prnewswire.com/news-releases/nagish raises-16-million-to-revolutionize-communication-for individuals-with-hearing-loss-through-ai-302192713
5. https://hearingreview.com/resourcecenter/webinars/beyond-audiograms revolutionizing hearing-aid-fitting-act-test
6. https://zeroproject.org/view/project/fbb37452-8e20-48ae-9068-04452c512781 International Journal of Advanced Research in Science, Communication an Technology 10.48175/ijarsct-2140 2021 pp. 212-216
7. HTML5 Working Draft, 24. 6. 2010, http://www.w3.org/TR/html5-diff/, accessed 5. 7.210
8. HTML5, A vocabulary and associated APIs for HTML and XHTML, W3C Working Draft,
9. Adobe Flash, http://www.adobe.com/flashplatform/, accessed 5. 7. 2010

4G 4G 15:28

Vo) 1 LTE2 25

From Swathi • swathi752005@gmail.com

To icmsciconference@gmail.com

Date Nov 13, 2024, 20:34

View security details

IYAL : REAL-TIME VOICE TO TEXT COMMUNICATION FOR THE DEAF

Mrs T Balamani, Assistant Professor, Department of Electronics and Communication Engineering, M Kumarasamy College of Engineering

Subiksha K, Swathi D, Vennila V, Vishnu Bharathi J, Pre-final year, Department of Electronics and Communication Engineering, M Kumarasamy College of Engineering

Abstract:

The Iyal app is a real-time communication tool designed to assist those who are hard of hearing or deaf in daily situations, much like the Nagish app. Through speech-to-text translation and vice versa, Iyal ensures seamless communication in both personal and professional contexts. Because the software is bilingual, users may communicate with ease regardless of the language being spoken. Its advanced AI speech recognition and text-to-speech algorithms produce accurate translations in real time. Reversible font sizes and background colors are among the adjustable accessibility options that Iyal provides to further meet various needs. This user-friendly program can assist the deaf population become more connected and

Real-time communication is one of the main benefits of voice-to-text phone call applications. These applications provide instantaneous, back-and-forth interactions, much like a typical phone call, in contrast to texting, which sometimes involves delays. This is especially helpful when answering customer service concerns, making arrangements, or setting up appointments—all tasks that require quick thinking and urgency.

Many deaf people have historically made phone calls using relay systems. These services, however, call for the involvement of a third party, which may cause privacy issues. Many people find it awkward to have someone else join them in conversation, whether they are talking about delicate subjects or they just want to engage in more intimate chat. Voice-to-text applications eliminate this barrier, enabling

PDF IEEE CONF...E PDF.pdf

2

Swathi Nov 18