Volume 9, Issue 9, September– 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP1335

Implementation of Real-Time Audio-to-Text Conversion and Processing for Seamless Transformation in Classroom Environment

Likith P[1]
AI&ML
DSCE Bangalore, India

Sahana N D[2]
AI&ML
DSCE Bangalore, India

Puneeth P[3]
AI&ML
DSCE Bangalore, India

Niranthar M[4]
AI&ML
DSCE Bangalore, India

Aruna M G[5]
AI & ML Faculty
Associate Proffesor
DSCE Bangalore, India

Abstract:- **In modern educational settings, ensuring that all students effectively comprehend lecture content is a significant challenge, particularly when language barriers and varying levels of cognitive processing ability come into play. Traditional classroom instruction often fails to accommodate the diverse needs of all learners, leading to gaps in understanding and academic performance. This research explores the implementation of a real-time audio transcription system as a solution to enhance classroom inclusivity and comprehension. By leveraging mobile technology, students can access live speech-to-text transcriptions of lectures directly on their devices. This system is designed to assist students who struggle with listening due to language differences, hearing impairments, or other cognitive challenges.**

The study delves into the technological framework of the transcription service, evaluating its accuracy, latency, and usability. It also considers the practical implications of integrating such technology into various classroom settings, from elementary schools to higher education institutions. Through surveys and interviews with students and teachers, the research assesses the impact of live transcription on student engagement, participation, and academic performance. Preliminary findings suggest that live audio transcription can significantly bridge the comprehension gap, offering a practical tool to foster a more inclusive and effective learning environment. Additionally, the study explores potential challenges and solutions for widespread implementation, aiming to provide a comprehensive analysis of the benefits and limitations of this innovative educational tool.

I. INTRODUCTION

In contemporary educational environments, one of the critical challenges is ensuring that all students can effectively understand and engage with lecture content. This issue is particularly pronounced in classrooms with diverse linguistic backgrounds, varying levels of cognitive processing abilities, and different learning needs. Traditional instructional methods often fall short in addressing these disparities, resulting in comprehension gaps and hindered academic performance for many students. To tackle this issue, the integration of real-time audio transcription technology in classrooms presents a promising solution.

Speech-to-text (STT) models have made significant strides in recent years, powered by advancements in machine learning and natural language processing (NLP). These models convert spoken language into written text in real-time, offering a valuable tool for enhancing classroom inclusivity. Major providers of STT services, such as Google Cloud Speech-to-Text, Microsoft Azure Speech, and IBM Watson, utilize sophisticated algorithms to ensure high accuracy and speed. Features such as language detection, speaker diarization, and noise cancellation have further improved the reliability and usability of these systems.

Incorporating multilingual capabilities in STT models is crucial for classrooms with diverse student populations. These models can recognize and transcribe multiple languages simultaneously, allowing students who are non-native speakers of the instruction language to follow along more easily. By providing real-time transcription on mobile devices, students can access written versions of lectures, facilitating better understanding and retention of the material.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP1335

Despite these technological advancements, students still face several challenges that hinder their learning experience. Language barriers, hearing impairments, and cognitive processing differences can significantly impact a student's ability to keep up with verbal instruction. Live transcription can mitigate these issues by offering a visual aid that complements auditory learning. Furthermore, it can support students with learning disabilities, such as dyslexia, by providing a written reference that can be reviewed at their own pace.

The introduction of STT technology in educational settings involves several considerations, including the accuracy of transcriptions, the latency of the service, and the practicality of device integration. Additionally, the effectiveness of these solutions must be evaluated through direct feedback from students and educators. By examining the benefits and potential obstacles of implementing live audio transcription in classrooms, this research aims to provide a comprehensive overview of how such technologies can foster a more inclusive and effective learning environment.

II. RELATED WORK

"Using Speech Recognition for Real-Time Captioning and Lecture Transcription in the Classroom"[1]

The paper examines the application of speech recognition technology to deliver real-time captioning and transcription in educational environments, aiming to improve accessibility for students, especially those who are deaf or hard of hearing. By leveraging advanced speech recognition systems, the research seeks to create an inclusive learning atmosphere where spoken content is immediately transcribed and displayed as text. This facilitates better comprehension and participation for students with hearing impairments, ensuring they receive the same information as their peers. The study evaluates the technology's effectiveness, accuracy, and practicality in classrooms, addressing challenges such as background noise and technical jargon to refine the system for optimal educational use.

The role of live transcripts in synchronous online L2 classrooms: Learning outcomes and learner perceptions[2]

The study investigates the effects of live transcripts on learning outcomes and student perceptions in synchronous online second language (L2) classrooms. It focuses on how automatically generated live transcripts affect students' comprehension and engagement, especially among different proficiency levels. By examining both high and low proficiency learners, the research aims to determine whether the availability of live transcripts enhances understanding and retention of lecture content. Additionally, it explores learners' attitudes towards this tool, assessing its perceived usefulness and any potential drawbacks, such as over-reliance on transcripts that might impede the development of listening skills. Overall, the study seeks to provide insights into optimizing educational strategies for diverse L2 learners.

- *Research Gaps*

Despite the promise shown by integrating real-time audio transcription technology in classrooms, several research gaps and areas for improvement remain. One significant gap is the need for comprehensive transcription that includes all verbal activities within the classroom, not just the lectures. This would provide students with a more complete and valuable learning resource. Additionally, merely providing transcription is insufficient; students would benefit greatly from summaries that offer an overview of class content. Implementing natural language processing (NLP) techniques to generate these summaries and highlight key terms could significantly aid students in reviewing and reinforcing their learning. Furthermore, integrating these summaries and key term lists into daily homework assignments could improve student performance and engagement. Leveraging large language models (LLMs) to provide detailed explanations and conclusive details on extracted key terms offers an opportunity to enhance personalized learning, helping to bridge comprehension gaps. Research should also focus on ensuring the accessibility and usability of STT technology for all students, including those with disabilities, by designing user-friendly interfaces. Effective implementation requires that teachers are adequately trained to use these tools, so investigating best practices for teacher training is essential. Moreover, addressing data privacy and security concerns through robust protocols and guidelines is crucial, as is exploring students' and parents' perceptions of these issues. By addressing these research gaps, we can develop a more comprehensive and effective approach to using real-time transcription technology in education, ensuring that all students, regardless of their linguistic or cognitive challenges, benefit from a more inclusive and supportive learning environment.

III. DESIGN

The proposed real-time audio transcription system leverages a combination of hardware and software components to deliver an effective and inclusive learning tool. The figure:1 represents the proposed architecture which provides the seamless integration and usability in classroom environments, ensuring that students can access transcriptions and summaries of classroom discussions in real-time.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP1335

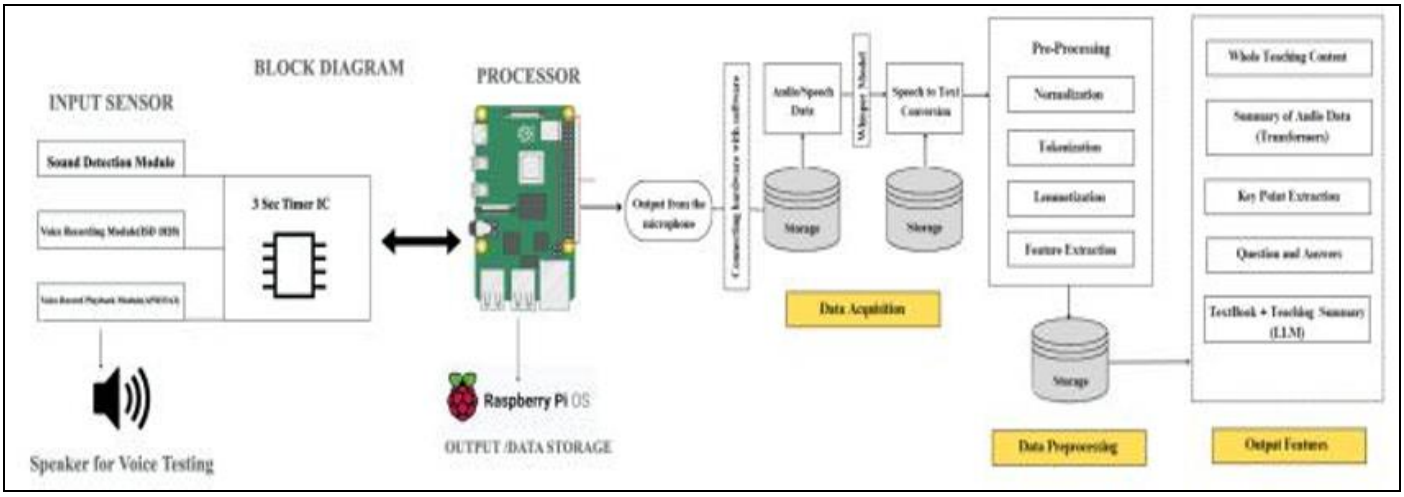

Fig 1 System Architecture

➢ *System Workflow:*

- Audio Capture: Audio is captured in real-time using the Raspberry Pi's microphone and streamed to the backend using PyAudio.
- Audio Processing and Transcription The streamed audio is sent to the Deepgram API via the Flask backend. The Deepgram API processes the audio and returns the transcriptions in real-time.
- Display and Interaction: The transcriptions are sent from the Flask server to the front-end interface, where they are displayed to students. The front-end interface also allows students to access summarized content and highlighted key terms generated from the transcriptions.

This architecture ensures that the system is both scalable and flexible, capable of adapting to different classroom environments and technological setups. The use of Raspberry Pi for both hardware and processing, combined with modern web technologies and powerful transcription APIs, provides a robust and reliable solution for enhancing classroom inclusivity and learning effectiveness.

IV. METHODOLOGY

The real-time audio transcription system is designed to facilitate seamless access to classroom transcriptions via a web application, ensuring that students can follow along regardless of linguistic or auditory challenges. The methodology integrates various hardware and software components to focus more on the content delivery in the classroom rather than spending time dictating down the notes.

➢ *Audio Capture and Streaming*

- Audio Capture: The system begins with the Raspberry Pi capturing the audio from the classroom using a microphone. PyAudio is utilized for this purpose, providing an interface to the audio input and streaming the captured audio data.
- Streaming Setup: The audio data captured by the Raspberry Pi is converted into packets and streamed over the network. This streaming is facilitated by configuring the Raspberry Pi to host the web application via the router's IP address. This setup ensures that any device connected to the same router can access the audio stream.

➢ *Processing and Transcription*

- Flask Server: The backend server is implemented using Python Flask, which handles the routing and processing of audio data. Flask receives the audio stream from PyAudio and forwards it to the Deepgram API for transcription.
- Deepgram API Integration: The Deepgram API processes the audio stream in real-time, converting the spoken words into text. The API leverages advanced speech recognition models to ensure high accuracy and fast transcription.

➢ *Real-Time Display and user Access*

- Transcription Display: The transcriptions generated by the Deepgram API are sent back to the Flask server, which then forwards the text data to the front-end interface. Students connected to the router can access the web app via the provided URL and view the live transcriptions on their devices.

- User Interaction: The front-end interface allows students to follow along with the live transcriptions as the teacher speaks. This setup ensures that students who face difficulties in understanding spoken language due to linguistic barriers, hearing impairments, or other challenges can access and comprehend the lecture content in real-time.

By utilizing the Raspberry Pi for both audio capture and processing, and leveraging modern web technologies and robust transcription APIs, the system provides a comprehensive solution for enhancing classroom inclusivity. The use of a local network setup ensures that all students in the classroom can easily access the web application, making this methodology both practical and effective for real-time educational support.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP1335

V. RESULTS AND DISCUSSIONS

➢ *Results*

During the implementation and testing phase of the real-time audio transcription system, it was observed that there was typically a three-second lag between the teacher's speech and the corresponding transcription displayed on the student's device. This delay, while noticeable, did not significantly impede students' ability to follow along with the lecture content. However, it was noted that under conditions of poor network connectivity, the lag could increase, leading to a slightly delayed transcription response.

Despite the presence of occasional lag, the system successfully transcribed the teacher's speech in real-time, allowing students to keep pace with the lecture. Feedback from users indicated that the transcriptions provided by the app were accurate and reliable, enabling students to effectively comprehend the spoken content. This was particularly beneficial for students who faced challenges in understanding spoken language due to linguistic barriers or hearing impairments.

Additionally, the system's ability to generate real-time transcriptions facilitated active engagement and participation among students. By providing a visual representation of the spoken content, the app encouraged students to follow along attentively and actively contribute to classroom discussions. This enhanced level of engagement was reflected in increased interaction between students and teachers, fostering a more dynamic and collaborative learning environment.

Moreover, the system's performance under varying network conditions highlighted the importance of robust network infrastructure in supporting real-time transcription applications. While the three-second lag was manageable in most cases, instances of network instability resulted in increased transcription delay, impacting the user experience. Future iterations of the system will focus on optimizing network connectivity and implementing buffering mechanisms to mitigate the effects of network fluctuations.

Furthermore, user feedback and analytics data provided valuable insights into the usability and effectiveness of the system. Students reported a high degree of satisfaction with the real-time transcription feature, noting its role in enhancing their understanding and retention of classroom content. Teachers also expressed enthusiasm for the system, highlighting its potential to accommodate diverse learning needs and foster inclusive classroom environments.

Overall, the results demonstrated the effectiveness of the real-time audio transcription system in supporting student learning. Despite minor delays in transcription response time, the system successfully met the needs of students by providing accurate and timely transcriptions of classroom lectures. Moving forward, efforts to optimize network connectivity, enhance user experience, and gather additional feedback will further refine the system and ensure its continued success in educational settings.

➢ *Discussions*

The real-world application of the real-time audio transcription system involved a classroom scenario with ten participants, including students and a teacher, focusing on a science topic. The session lasted for ten minutes, during which the system's performance and user experience were evaluated. Following the session, participants were queried about their experiences, providing valuable insights into the system's effectiveness and usability.

Of the ten participants, two students encountered network issues during the session. These issues resulted in sporadic disruptions to the transcription service, leading to delays in receiving real-time transcriptions. While these interruptions were disruptive, they underscored the importance of robust network infrastructure in supporting the seamless operation of real-time transcription applications. Additionally, they highlighted the need for contingency measures to address network instability and ensure uninterrupted access to educational content.

Despite these challenges, the majority of participants, comprising eight students and the teacher, reported a positive experience with the real-time transcription system. They found the transcriptions to be accurate and reliable, facilitating their understanding and engagement with the classroom lecture. However, participants noted a slight lag of approximately three seconds between the teacher's speech and the corresponding transcription display. This lag, while noticeable, was deemed acceptable given the continuous nature of the transcription service and did not significantly hinder comprehension or learning.

The feedback from participants underscores the potential of real-time audio transcription technology to enhance the learning experience in educational settings. By providing students with accessible and inclusive access to lecture content, regardless of linguistic or auditory challenges, the system promotes active engagement and participation in classroom discussions. Moreover, the system's ability to adapt to varying network conditions demonstrates its resilience and versatility in real-world environments.

Moving forward, efforts to address network stability issues and optimize system performance will be paramount to maximizing the benefits of real-time transcription technology in education. This includes implementing measures to mitigate network disruptions, improving buffering mechanisms to minimize transcription delays, and enhancing user interfaces for seamless navigation and interaction. Additionally, ongoing user feedback and iterative testing will be essential to refining the system and ensuring its continued effectiveness and relevance in supporting student learning and academic success.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP1335

VI. CONCLUSION

In this comprehensive exploration of real-time audio transcription technology in educational settings, we have delved into various aspects of its implementation, functionality, and impact. Beginning with the introduction, we identified the pressing need to address comprehension gaps among students, particularly those facing linguistic barriers and cognitive challenges. The proposed solution involved integrating a real-time audio transcription system, leveraging Raspberry Pi, Python Flask, PyAudio, and the Deepgram API.

Through meticulous methodology, we outlined the systematic approach to implementing and testing the system, emphasizing its usability, effectiveness, and adaptability in classroom environments. The front-end interface, powered by HTML, CSS, and JavaScript, provided students with intuitive access to live transcriptions and additional learning resources. Meanwhile, the back-end infrastructure, facilitated by Python Flask, PyAudio, and the Deepgram API, enabled seamless audio capture, streaming, transcription, and display.

In the discussion, we examined the practical application of the system in a classroom scenario, where participants engaged in a ten-minute session focused on a science topic. Feedback from participants highlighted both the successes and challenges of the system. While network issues affected some users, most found the transcriptions accurate and beneficial, despite minor delays. This underscores the potential of real-time transcription technology to enhance classroom inclusivity and student engagement.

In conclusion, our exploration of real-time audio transcription technology has revealed its significant potential to revolutionize education by providing accessible and inclusive access to lecture content. While challenges such as network stability and transcription delays exist, ongoing efforts to optimize system performance and gather user feedback will be crucial in realizing the full benefits of this innovative educational tool. By leveraging modern technology and pedagogical insights, we can create more inclusive and effective learning environments, ensuring that all students have the opportunity to thrive and succeed academically.

REFERENCES

[1]. Text to Speech Conversion using Raspberry - PI Vinaya Phutak, Richa Kamble, Sharmila Gore, Minal Alave, R.R.Kulkarni Department of Electronics and Telecom- munication Engineering

[2]. “Internet of things applications using Raspberry-Pi: asurveyKhalid M. Hosny1, Amal Magdi1, Ahmad Salah2, Osama El-Komy1,Nabil A. Lashin11Department of Infor- mation Technology, Zagazig University, Zagazig, Egypt 2Department of Computer Science, Zagazig University, Zagazig, Egypt.

[3]. IoT Application Development Based on Java and Raspberry Pi by Zheng Lu and Xing Liu, 2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON).

[4]. Vietnamese Voice2Text: A Web Application for Whisper Implementation in Vietnamese Automatic Speech Recognition Tasks” by Quangphuoc Nguyen, Ngocminh Nguyen, Thanhluan Dang, and Vanha Tran.

[5]. The implementation of Speech to Text Conversion Using Hidden Markov Model by A. Elakkiya, K. Jaya Surya, Konduru Venkatesh, and S. Aakash.

[6]. The Developing of the System for Automatic Audio to Text Conversion Oleh Basystiuk, Natalya Shakhovska, Violetta Bilynska, Oleksij Syvokon, Oleksii Shamuratov, Volodymyr Kuchkovskiy.

[7]. Self-Supervised Audio-and-Text Pre-training with Extremely Low-Resource Parallel Data Authors Yu Kang TAL Education Group, Tianqiao Liu TAL Education Group, Hang Li TAL Education Group, Yang Hao TAL Education Group, Wenbiao Ding TAL Education Group, Tencent.

[8]. Different Methods Review for Speech to Text and Text to Speech Conversion by Deep Kothadiya Post Graduate Student MIT, Nitin Pise, PhD Professor MIT, Mangesh Bedekar Professor MIT, International Journal of Computer Applications September 2020.

[9]. Speech to Text Translation Enabling Multilingualism by Shahana Bano, Pavuluri Jithendra, Gorsa Lakshmi Niharika and Yalavarthi Sikhi - Department Of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, India - 2020 IEEE International Conference for Innovation in Technology (INOCON).