e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

REAL TIME SUBTITLE GENERATION FOR VIDEO CONFERENCING

Prof. Runal Pawar[*1], Vaishanv Ghenge[*2], Jayesh Nikumbh[*3],
Atharva Nigal[*4], Atharva Tapkir[*5]

[*1]Assistant Professor, Department Of Computer Engineering, Sinhgad College Of Engineering, Pune, India.

[*2,3,4,5]Department Of Computer Engineering, Sinhgad College Of Engineering, Pune, India.

DOI : https://www.doi.org/10.56726/IRJMETS51897

ABSTRACT

The Video conferencing has emerged as a crucial tool for remote communication and collaboration, facilitating interactions among individuals and organizations across geographical boundaries. However, traditional video conferencing systems often overlook the needs of users with hearing impairments or those operating in noisy environments, leading to accessibility challenges and communication barriers. Real-time subtitle generation presents a promising solution to address these challenges by providing synchronized text captions alongside the video feed, enhancing accessibility and comprehension for all participants. In this paper, we propose a novel approach to real-time subtitle generation for video conferencing applications. Our system leverages automatic speech recognition (ASR) technology to transcribe spoken dialogue into text, which is then synchronized with the corresponding video frames in real-time. The system architecture consists of three main components: audio processing, speech recognition, and subtitle generation. The audio processing module preprocesses the incoming audio stream to remove noise and enhance speech clarity, while the speech recognition module utilizes state-of-the-art deep learning techniques to accurately transcribe the processed audio into text. Finally, the subtitle generation module synchronizes the transcribed text with the video frames to generate real-time subtitles. Overall, our research contributes to the advancement of real-time subtitle generation technology and its integration with video conferencing platforms. By seamlessly integrating ASR with video conferencing systems, our approach offers an efficient and effective solution to enhance accessibility and comprehension for users with hearing impairments or those operating in noisy environments. Future work may involve further optimization of the system for improved accuracy and scalability, as well as exploration of additional features such as multi-language support and speaker identification.

Keywords: Real-time Subtitle Generation, Video Conferencing, Automatic Speech Recognition, Accessibility, Virtual Communication.

I. INTRODUCTION

Video conferencing has become an integral part of modern communication, facilitating remote collaboration and interaction across geographical boundaries. However, traditional video conferencing systems often overlook the needs of users with hearing impairments or those in noisy environments where audio clarity is compromised. Real-time subtitle generation offers a solution to address these challenges by providing synchronized text captions alongside the video feed, enhancing accessibility and comprehension for all participants. In this paper, we present a novel approach to real-time subtitle generation for video conferencing applications. Our system leverages automatic speech recognition (ASR) technology to transcribe spoken dialogue into text, which is then synchronized with the corresponding video frames in real-time. By seamlessly integrating ASR with video conferencing platforms, our system offers an efficient and effective solution to enhance communication accessibility for users with hearing impairments and those operating in noisy environments.

II. LITERATURE SURVEY

The literature on real-time subtitle generation and automatic speech recognition (ASR) encompasses a wide range of approaches, methodologies, and applications. In this section, we review key studies and advancements in these areas, providing insights into the current state-of-the-art and identifying relevant research trends.

2.1 Real-time Subtitle Generation:

Real-time subtitle generation has gained significant attention in recent years due to its potential to enhance

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

accessibility and comprehension in various multimedia applications. Chiu et al. (2017) proposed a deep learning-based approach for real-time subtitle generation in multimedia content, demonstrating the effectiveness of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) in transcribing spoken dialogue into text. Their study highlighted the importance of leveraging deep learning techniques for improving the accuracy and efficiency of real-time subtitle generation systems.

Xu et al. (2019) conducted a comprehensive survey on automatic speech recognition systems, providing an overview of different methodologies, algorithms, and evaluation metrics. The survey identified key challenges in ASR, such as vocabulary coverage, noise robustness, and speaker variability, and discussed potential solutions and future research directions. Their findings underscored the importance of continuous advancements in ASR technology to support real-time subtitle generation in diverse applications.

2.2 Automatic Speech Recognition (ASR):

Advances in deep learning have significantly contributed to the progress of automatic speech recognition systems. Graves et al. (2013) introduced deep recurrent neural networks (RNNs) for speech recognition, demonstrating superior performance compared to traditional hidden Markov models (HMMs). Their work laid the foundation for the development of deep learning-based ASR systems, which have since become the standard approach in the field.

Hori et al. (2017) provided an overview of recent advancements in deep learning for audio signal processing, highlighting the role of convolutional neural networks (CNNs), recurrent neural networks (RNNs), and attention mechanisms in improving the accuracy and robustness of ASR systems. Their review emphasized the importance of exploring novel architectures and algorithms to address challenges such as speaker variability, background noise, and domain adaptation.

2.3 Integration of ASR with Video Conferencing:

The integration of ASR technology with video conferencing platforms offers promising opportunities to enhance communication accessibility and inclusivity. Several commercial platforms, such as Zoom and Microsoft Teams, have incorporated real-time transcription features to support users with hearing impairments and facilitate multilingual communication. These advancements underscore the growing importance of real-time subtitle generation in modern communication systems.

RELATED WORK

Previous research has explored various approaches to real-time subtitle generation, including rule-based methods, statistical models, and deep learning techniques. While these approaches have shown promising results, they often suffer from limitations such as high computational complexity, limited vocabulary coverage, and difficulty in handling noisy audio input.

III. SOFTWARE AND HARDWARE REQUIREMENTS

Model and Material which are used is presented in this section. Table and model should be in prescribed format.

- **SOFTWARE**

Sr. No.	Software	Description
1	NextJS, WebRTC	Framework used for building Front-End development part and Open Source Protocol used for real time communication.
2	NodeJS, ExpressJS	Framework used for Backend-End development part.
3	RESTful Web Services, Spring Boot Security, OAuth2.0, JSON Web Tokens, PostgreSQL	API's, Authentication and Verification, and database for storage.

- **HARDWARE**

Sr. No.	Software	Description
1	Server Hardware	
1.1	CPU	A multi-core processor (e.g., quad-core or higher) to handle concurrent requests efficiently.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

1.2	RAM	At least 8GB of RAM.
1.3	Storage	A SSD with ample storage space.
2	Client Hardware	
2.1	CPU	A multi-core processor(dual core or higher).
2.2	RAM	A minimum of 2GB of RAM for smooth operation
2.3	Storage	At least 16GB is recommended.
2.4	Capacitive Sensor	Web Cam and microphone.

IV. CONCLUSION

In conclusion, we have presented a novel approach to real-time subtitle generation for video conferencing applications, leveraging ASR technology to provide synchronized text captions alongside the video feed. Our system offers an effective and efficient solution to enhance accessibility and comprehension for users with hearing impairments or those in noisy environments. Future work may involve further optimization of the system for improved accuracy and scalability, as well as exploration of additional features such as multi-language support and speaker identification.

V. FUTURE SCOPE

- **Machine Learning for Fraud Detection:** Implement machine learning algorithms to detect and prevent fraudulent activities during the voting process. This could include anomaly detection based on voting patterns or behavior analysis to identify potential threats.
- **Usability and User Experience Enhancements:** Focus on improving the usability and user experience of the voting system. Conduct user studies to identify potential pain points and implement design changes to make the system more user-friendly and accessible to a diverse population.
- **Open Source Collaboration:** Consider making the voting system an open-source project, allowing collaboration with the wider community. This can lead to continuous improvements, peer reviews, and a more transparent development process.
- **Cybersecurity Audits and Continuous Monitoring:** Implement regular cybersecurity audits and continuous monitoring to identify and address potential vulnerabilities promptly. This proactive approach will help in maintaining the system's resilience against evolving cyber threats.
- **International Standards Compliance:** Ensure that the voting system complies with international standards for secure electronic voting. This may involve collaboration with international organizations and adherence to established guidelines to promote trust and reliability on a global scale.

VI. REFERENCES

[1] Ganesh Prabhu S, Prabu.S, R.R. Thirruanvukkarasu, Nizarahammed.A, Raghul.S,P Jayarajan., “Smart Online Voting System”,IEEE,2021.

[2] Jehovah Jireh Arputhamon, Dr.A.Gnana Saravanan.,“Online Smart Voting System Using Biometrics Based Facial and Fingerprint Detection on Image Processing and CNN”,IEEE,2021.

[3] Suryaa Pranav Meduri, Anupama Meduri, Saketh Kamatham,Neha Diwan, Sneha Subramanian,,“ A Secure Network Monitored Balloting System”,IEEE,2021.

[4] Mr. Pradeep kumara V. H, Dr. Ravindra P. Rajput, “Design and implementation of convenient and compulsory voting system using finger print sensor and GSM technologies.”,IEEE,2020.

[5] Dinesh Kumar, Akshay Hareendran, Askar Ali, Bharanidharan.,“ Online Voting System Using Aadhaar Card and Biometric”, ISSN,2020.