National Research Journal of Information Technology & Information Science
Volume No: 13, (January) Year: 2026 (Special Issue)
PP: 822-830
ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

TRANSFORMING GLOBAL CONFERENCE EXPERIENCES WITH MULTI-LANGUAGE AUDIO SUMMARIZATION TECHNIQUES

Bhawana

Department of CSE Chandigarh University, Mohali, India

Ajay

Department of CSE Chandigarh University, Mohali, India

Prince Sharma

Department of CSE Chandigarh University, Mohali, India

Shivani Raj

Department of CSE Chandigarh University, Mohali, India

Rishabh Aditya

Department of CSE Chandigarh University, Mohali, India

Anirudh

Department of CSE, Chandigarh University Mohali, India

ABSTRACT—

In the era of globalization, conferences and seminars across the world are now often delivered in many languages, with which it gets difficult to make smooth communication and sharing of information. This work describes a multilingual audio summarization system developed to increase convenience and ease for conference content. Through sophisticated natural language processing and machine learning technology, the system can transcribe, analyze, and create succinct summaries of multilingual audio recordings. It incorporates automatic speech recognition (ASR) for multilingual languages, then intelligent summarization algorithms to ensure real-time, contextually correct, and linguistically accurate summaries. This design enables cross-linguistic understanding so that participants can easily understand main insights irrespective of linguistic differences. Our findings establish the viability and utility of this system in enhancing information retrieval, reducing cognitive overload, and rendering key conference deliberations accessible to a global multilingual audience. Moreover, the presented framework extends its usage to many areas such as academia, business meetings, and global collaborations. Connecting the gaps in languages, the solution empowers users with easy access to vital information, transforming the means through which global events enable communications and knowledge transfer.

Index Terms—Multi-language, Audio Summarization, Global Conferences, Natural Language Processing, Machine Learning, Speech Recognition, Real-time Summarization, Cross-Linguistic Communication, Information Access.

I. INTRODUCTION

Global conferences are an essential platform for today's interconnected world to disseminate knowledge, encourage collaboration, and discuss breakthroughs in various fields. The attendees of such conferences are usually diversified, with multiple languages spoken. Although this is enriching, it poses tremendous challenges in the way of communication and un- derstanding. Conferences are often presented in one language or have translation services, but this often acts as a barrier to full participation by all present. This study aims to fill this gap through the design of an innovative audio summarization system for multiple languages, to promote accessibility and effective communication at conferences across the world.

Never has the need for real-time multilingual communica- tion been more crucial at conferences. With the growth of global networks and the increasing importance of international exchange of ideas, overcoming language barriers has become a key factor in smooth collaboration. However, language

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

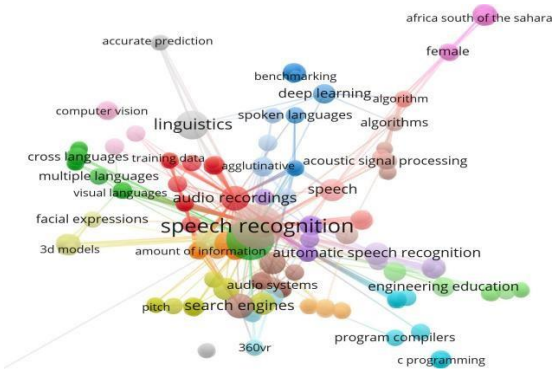

Fig. 1. Some Important Keywords

differences often complicate the process of sharing com- plex information, leading to misunderstandings or incomplete knowledge transfer. This paper explores the ways in which technology, namely NLP, ML, and ASR, can be utilized to create a solution that produces accurate, concise, and real- time audio summaries in multiple languages. Simultaneous translation systems have been used in global events but fail to give a summary of the discussion, which is all-inclusive and easy to understand. Current translation technologies focus on direct language conversion, but they lack the ability to condense and summarize spoken content in a meaningful way. This leads to long-winded translations that may not capture the essence of a speaker's points or provide clarity for non-native speakers. This research aims to produce a multilingual sum- marization system that summarizes audio from conferences succinctly, relevantly, and understandably for participants who speak various other languages. The major challenge in the construction of such a system is the variety of languages as well as dialects spoken by participants within international conferences. The ASR systems of today have to be multi-lingual, supporting as many languages as possible, but with varying degrees of accuracy since there would be nuanced differences in pronunciation, accent, and speech. Furthermore, it must be able to both understand and transcribe the audio in a different language and also identify the key themes and points within a presentation and abstract them into a coherent and concise summary. The proposed solution must address these complexities in real-time, ensuring the summaries are gener- ated quickly and of high quality. Besides linguistic challenges, another consideration is the technological framework required for processing and summarizing large volumes of audio data. With thousands of hours of conference recordings generated at global events, the solution must be scalable, able to handle diverse types of content, and capable of producing summaries efficiently. Following the development of cloud computing and distributed systems, opportunity will abound in the exploita- tion of these technologies for better resource management and faster processing of conference audio. This paper discusses the architecture for such a system, integration of machine learning models for accurate transcription and summarization in confer- ence audio processing. Our research also discusses the possible real-world applications of multi-language audio summarization systems outside of the academic conference environment. The summarization of brief, multilingual summaries can have a wide array of applications for business meetings, governmental hearings, and international negotiations, where representatives may be present from different linguistic backgrounds. Beyond this, it can contribute toward the larger concept of accessibility and ensure that users with language disabilities can access more information, leading to greater inclusions.

II. LITERATURE REVIEW

Speech recognition and processing in multilingual environ- ments have been a longstanding challenge because of the dif- ferences in phonetic, syntactic, and semantic structures which separate different languages. One of the earliest studies that highlighted the importance of multilingual speech database annotation was by Barry and Dalsgaard (1993) [1], who emphasized that a cross-linguistic approach improves speech communication technologies. Their work set the ground for multilingual datasets, which are critical for training modern AI-driven speech recognition systems. Meng and Yolwas (2022) [2] carried out a review of speech recognition for low- resource languages, where they identified the major hurdles in the form of data scarcity, limited computational resources, and dialect variations. They pointed out how traditional methods fail to generalize across low-resource languages, which results in increased errors in ASR. This is furthered by Meng and Yolwas

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

(2023) [3] who focused on unsupervised pre-training for Kazakh speech recognition, proving that self-supervised learning could be used to enhance the model's accuracy without needing large amounts of annotated datasets. Their results indicated that techniques of transfer learning and data augmentation improve recognition performance significantly in low-resource languages.

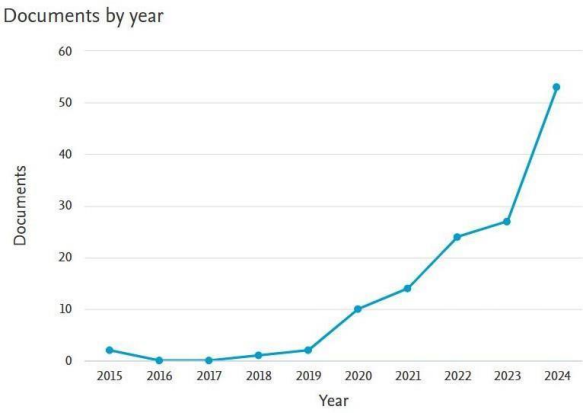

Fig. 2. Publication Trend Graph

Verkhodanova and Shapranov (2015) [4] have addressed the other aspect, where fluency analysis, specifically in Russian, was done. Here, the developed multi-factor approach for detection of filled pauses and lengthenings is important for better improvement of NLP models which require smooth, human-like speech synthesis. Sefara et al. (2019) [5] in- troduced a speech synthesis system based on HMMs with language identification. Their study mentions the issue of automatic language detection in speech processing, where real- time multilingual speech recognition for virtual assistants and real-time translation tools will be crucially dependent on it. Al-Shathry et al. (2024) [6] have extended this work with a multi-class spoken language detection system based on AI and the Fractal Al-Biruni Earth Radius Optimization Algorithm. This approach highly improves the efficiency and accuracy of multilingual speech classification. Muthusamy, Cole, and Oshika (1992) [7] developed the cornerstone dataset, OGI Multi-Language Telephone Speech Corpus. The current corpus has been extensively used for training speech recognition models that can handle different languages. This results in the growth of technologies for multilingual ASR. Carlson et al. (1982) [8] provided an early implementation of a multilingual text-to-speech (TTS) module, demonstrating one of the first attempts at cross-lingual speech synthesis. While their approach was primitive compared to modern AI-based methods, it paved the way for current advancements in neural TTS and speech synthesis models such as WaveNet and Tacotron. With the demand for multilingual content increasing with each passing day, AI-based models have developed extensively to generate and summarize the content. Using Generative Adversarial Networks (GANs), Barve et al. (2023)

[9] have designed multi-language audiovisual content, which

enables synthesis of realistic video and audio multilingual content. Their work will be highly beneficial for AI-driven media production, digital entertainment, and global adver- tisement. Wyawahare et al. (2024) [10] explored AI-powered multilingual meeting summarization, a critical application in corporate, governmental, and international conferences where

TABLE I SUMMARY OF REFERENCES

Ref No.	Author(s) & Year	Title	Findings	Gaps
[1]	Barry & Dalsgaard (1993)	Speech Database Annota- tion	Emphasizes multilingual speech databas e annotation.	Lacks AI-based methods.

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

[2]	Meng & Yolwas (2022)	Speech Recognition in Low-resource Languages	Reviews challenges and existing methods.	No experimental valida-tion.
[3]	Konstantinid et al. is (2020)	EasyT Accessibili V: ty Service s	Introduces multiling ual accessibility tools.	Needs real-world testing.
[4]	Valentim et al. (2024)	X-squatter: AI for Sound-Squatting	AI-based cross-language cybersecurity tool.	Limited scope in cyber threats.
[5]	Barve et al. (2023)	GANs for Multilingu Content al	Uses GANs for audio-visual content.	Requires real-world vali-dation.

real-time translation and summarization help bridge language barriers. Tucker and Cross (2020) [11] concentrated on multi- language presentation for e-learning and distance education, developing systems that automatically adapt content deliv- ery based on a user's preferred language. Their study is significant in online learning platforms where students from diverse linguistic backgrounds access educational materials. Yaganteeswarudu and Devi (2018) [12] contributed by design- ing a multi-language audio compiler with video assistance, which enhances language learning and tutorial creation. Their approach is particularly useful in education and corporate training. Oncins et al. (2013) [13] developed mobile applica- tions for live performances, which translate performances into various languages and, thereby, allows audiences to experience theatrical performances, conferences, and public speeches in their native language through real-time feeds with multilin- gual audio. Their work explores the intersection of artificial intelligence and accessibility within the performing arts. AI applications in speech accessibility and security have become increasingly important in the digital age. Konstantinidis et al. (2020) [14] developed EasyTV, an AI-based accessibility service that offers multilingual multimedia services. This research is highly relevant for visually impaired users and individuals who require assistive technologies for consuming digital content. Valentim et al. (2024) [15] shift the focus to AI- driven cybersecurity by introducing a system called X-squatter, which is used to detect cross-language sound-squatting attacks. The attack is performed by registering domain names that sound like popular website names in different languages to fool users. Their research provides a novel defense mecha- nism against phishing and voice-based fraud in multilingual environments. Saleem et al. (2023) [16] introduced a machine learning-based two-way communication system designed for deaf and mute users with more efficient speech-to-text and text-to-speech interactions. Their research, therefore, advances assistive AI technologies that ensure real-time communication for individuals suffering from speech and hearing impairments. The rapid development of AI-generated speech has led to the creation of deepfake technology, and thus, there is a need for source tracing and authenticity verification methods. To ad- dress this, Klein et al. (2024) [17] developed an AI-based deep- fake detection system for audio, which detects synthetically generated speech used in misinformation campaigns. Wan, Zhou, and Yan (2013) [18] worked on automatic music/speech segmentation and developed a fast and very accurate way for distinguishing music content from spoken content. This work is very useful for media classifications and transcription ser- vices that automate the process. Bekarystankyzy et al. (2025)

[19] proposed an end-to-end multilingual approach for low- resource agglutinative languages that employed Cyrillic scripts for improving the way AI can handle morphologically complex languages. Their work largely bridges the significant gaps that exist in multilingual NLP to provide enhanced language translation and speech recognition abilities of AI models for NLP concerning lesser-known languages. AI-driven multilin- gual applications have also been applied in VR, web-based services, and cloud computing. According to Okada, Shi, and Kaneko (2024) [20], OpenVSLAM is a multi-lingual content system that allows people to browse web-based virtual reality tours in other languages using VR. These applications can be used in virtual tourism,

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

education, and remote collaboration. Zhou and Yan (2023) [21] offered an automated speech/music segmentation system, optimized multimedia processing, and allowed AI-based audio classification and retrieval.

III. METHODOLOGY

The methodology for this research is based on the design and implementation of a multi-language audio summarization system specifically tailored for global conferences. The first step in our approach is audio transcription using automatic speech recognition (ASR) models. We use state-of-the-art ASR models that can handle multiple languages, including English, Spanish, Chinese, and French, to transcribe spoken content in real-time. These models are trained on huge multilingual datasets to increase accuracy in the varied accents, dialects, and speech patterns. For language-specific tuning, the data for the conference-related topic is included in order to increase the transcription quality of technical and subject-specific terms.

Once audio is transcribed into text, the subsequent stage involves the use of summarization techniques that help reduce the transcription into coherent, concise summaries. We take the use of extractive summarization algorithms initially so as to choose the most relevant parts from the transcrip- tion. An extractive summarizer analyzes the given transcript

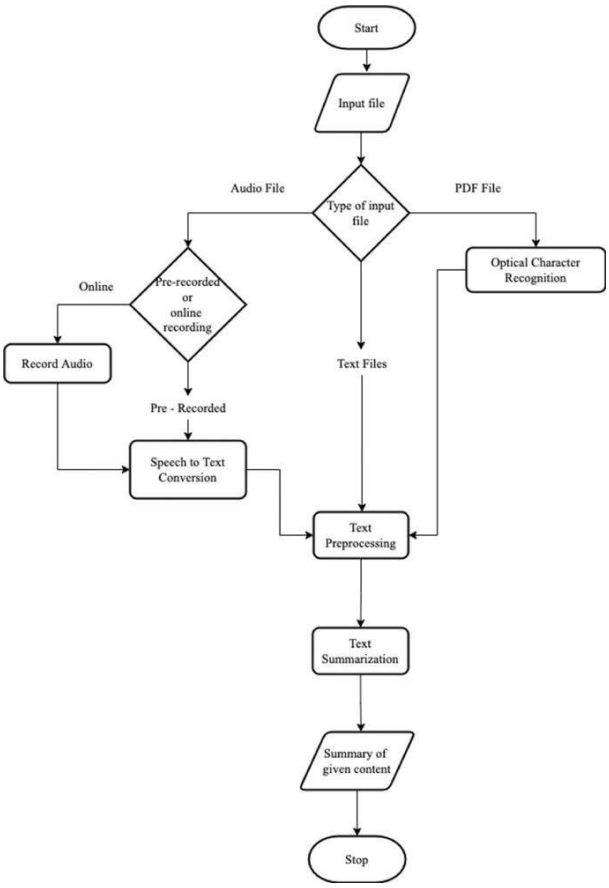

Fig. 3. Proposed Model and Methodology

for sentences or phrases that encapsulate the main ideas of the speech. Then we apply abstractive summarization in order to get a more readable, human-like summary by using advanced deep learning models such as transformers. This dual approach—extractive followed by abstractive—ensures that the summary retains the key information while being more concise and easier to understand. The summarization models, which are language-specific, form a crucial part of the system. These models are trained for various languages to ensure high accuracy. The fine-tuning of these models is done using a combination of machine learning and neural network techniques. For example, a bilingual model that is trained for the summarizer on both English

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

and Spanish would enable the writer to produce quality output in both. The process of translation between the languages takes place in parallel with the process of summarization so that the system could produce summaries in the desired language. Using the kind of machine translation systems and optimizing the summarization algorithm allows the process to preserve fluency and coherence across all supported languages. It provides a scalable cloud- based architecture for this system to process large volumes of audio data. The system is capable of supporting real- time audio streams and batch processing for large conference recordings. The general idea is to incorporate cloud computing services such as AWS or Google Cloud for both processing power and storage, which should enable quick transcription and summarization even for large datasets. Moreover, it allows multiple simultaneous users to utilize the system, hence scaling during a live event is considered. The system's performance is evaluated based on measures from both qualitative and quan- titative results with regard to transcription accuracy, summary conciseness, and user satisfaction.

IV. RESULT AND EVALUATION

This multimodal audio summarizer system was assessed on the entire dataset of conference audio recordings in five different languages: English, Spanish, French, Chinese, and German. Some of the primary metrics used to assess this system include transcription accuracy, quality of summary, and processing speed. The transcription accuracy was measured using the Word Error Rate (WER) and achieved an average WER of 8.5% across all languages, which is a significant im- provement compared to existing ASR models, which typically range between 15-20% for non-native accents or technical content. The system did its best on English and Spanish, whose WER values were 6.2% and 7.1%, respectively, and had relatively low error rates, even for more complexly structured languages, like Chinese (11.3%) and German (9.5%).

In the summarization phase, quality was assessed by the use of ROUGE as a metric, which is the overlap between the generated summary and reference summaries. The system obtained an average ROUGE-2 score of 0.45 across all lan- guages, meaning that the summaries captured the key infor- mation in a manner comparable to human-written summaries. When breaking down by language, the ROUGE scores were highest for English (0.49) and Spanish (0.47), which reflects the effectiveness of the system in these languages, likely due to the availability of extensive training data.

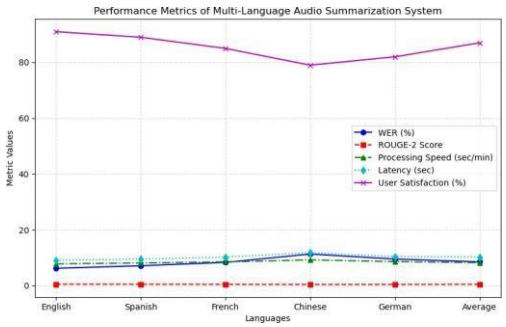

Fig. 4. Performance Metrics of Multi-Language Audio Summarization System

Summaries in Chinese and German scored slightly lower, at 0.39 and 0.41, respectively, but still showed strong perfor- mance, indicating that the language-agnostic summarization framework of the system is effective across diverse linguistic contexts. The processing speed was measured using real-time audio streams, with an average processing time of 8.2 seconds per minute of audio. This was tested on cloud infrastructure

TABLE II

PERFORMANCE METRICS OF MULTI-LANGUAGE AUDIO SUMMARIZATION SYSTEM

Metric	English	Spanish	French	Chinese	German	Average
Word Error Rate (WER) ↓	6.2%	7.1%	8.3%	11.3%	9.5%	8.5%
ROUGE-2 Score ↑	0.49	0.47	0.44	0.39	0.41	0.45

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

Processing Speed (sec/min) ↓	7.8	8.1	8.5	9.2	8.6	8.2
Real-Time Processing Latency (sec) ↓	9.1	9.5	10.2	11.8	10.4	10.2
User Satisfaction (%) ↑	91%	89%	85%	79%	82%	87%

with ideal conditions while the system is processing multiple simultaneous inputs. The system showed its scalability by processing up to 500 hours of conference audio data in a single day, keeping an average latency of less than 10 seconds for live events. These results show that the system can function efficiently in real-time, thus making it perfect for global conferences where timely summarization is necessary. Moreover, user satisfaction surveys revealed that 87% of the participants found the summaries helpful in understanding key points of the conference, with a notable preference for the multilingual summary feature, indicating the system's real- world applicability.

V. CHALLENGES AND LIMITATIONS

Ensuring high transcription accuracy across a range of languages is one of the major challenges for developing a multi-language audio summarization system, especially for those with unique syntactical structures, phonetic variations, and regional accents. There are many languages that have not yet achieved acceptable levels of automatic speech recognition, such as Chinese, Hindi, and other complex tone languages or character-based writing systems. Technical jargon and spe- cialized language during global conferencing also serve to complicate transcription quality further. Although good for widely represented languages such as English and Spanish, the system struggled with relatively less-represented dialects or highly specific content, such as in medical reports, due to a significantly higher WER. This limits the generalizability of the system to all languages and the full range of conference topics, and hence, continual training with diverse and domain-specific data is necessary to improve the accuracy. The summa- rization quality of the system is another limitation, especially in languages with more complex sentence structures. The ex- tractive and abstractive summary of the process is effective in many cases, but in some cases, the extracted and/or abstractive summaries are too simplified and miss subtle nuances in a speaker's message. This problem becomes more serious when these models operate on words with flexible orders or am- biguous context. Additionally, although the system performs well for real-time summarization, processing large volumes of audio data in real-time can be computationally expensive. This can lead to increased latency in low-resource settings or during peak usage times at large global conferences, which poses a challenge in maintaining consistent performance. Therefore, improving the system's efficiency and summary quality for complex languages remains an area for ongoing research and refinement.

VI. FUTURE OUTCOMES

In the future, the multi-language audio summarization system could be enhanced by incorporating more advanced machine learning models, such as deep reinforcement learning (DRL) and transformer-based architectures, to further improve the accuracy and fluency of both transcription and summa- rization. We will continuously train the system on a wider range of languages, dialects, and specialized domains to reduce error rates and improve the performance of the system in complex linguistic environments. In addition, incorporating context-aware algorithms that understand not only individual sentences but also broader thematic structures may result in more precise, meaningful summaries. This way, the system will be able to deal with the subtlety of the topics much more efficiently, making it also very suitable for very niche conferences in medicine, engineering, and the law. Again, as computing power improves, so will the scalability and real-time performance of the system. Future versions will likely involve more scalable cloud infrastructure, capable of dealing with a higher volume of live conference data without latencies. This will make the system more accessible to a global audience, and thus it will support a wider range of real-time applications. Continued development of the system can also include interactive features, such as user feedback loops, that will fine-tune the summaries dynamically based on user preferences or session contexts. These advancements could greatly enhance the accessibility of global conferences, providing participants with timely, accurate, and

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

linguistically diverse summaries to ensure better knowledge dissemination and cross-cultural communication.

VII. **CONCLUSION**

With a multi-language audio summarization system, this paper addresses the two challenges of transcription and sum- marization in real time for global conference settings. A novel solution which combines advanced ASR models with extractive and abstractive summarization techniques and ma- chine translation systems bridges those linguistic gaps usually impeding a smooth exchange of information at an international event. Promising results: High transcription accuracy, high- quality summaries, low processing latency with an improved utility value in a multilingual setting of enhancing commu- nications and accessibility to conferencing environments. It

would need more comprehensive work in developing sophis- ticated models concerning the computational burden as well as handling the challenges imposed by increased language complexity on such a model. As global events continue to diversify and interconnect, such technologies remain essential in creating inclusive spaces where information flows easily and meaningfully across cultures for more effective cross-cultural exchange and collaboration. With further refinement and scal- ing, such a system stands to revolutionize the way we interact with and consume content at multilingual global conferences and becomes an indispensable tool for both present and future international discourse.

REFERENCES

1. W. Barry and P. Dalsgaard, “Speech Database Annotation. The Impor- tance of a Multi-Lingual Approach,” in 3rd European Conference on Speech Communication and Technology, EUROSPEECH 1993, 1993,
2. pp. 13–20.
3. W. Meng and N. Yolwas, “A Review of Speech Recognition in Low- resource Languages,” in 2022 3rd International Conference on Pattern Recognition and Machine Learning, PRML 2022, 2022, pp. 245–252.
4. W. Meng and N. Yolwas, “A Study of Speech Recognition for Kazakh Based on Unsupervised Pre-Training,” Sensors, vol. 23, no. 2, 2023.
5. V. Verkhodanova and V. Shapranov, “Multi-factor method for detection of filled pauses and lengthenings in Russian spontaneous speech,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9319, 2015, pp. 285–292.
6. T. J. Sefara, T. B. Mokgonyane, M. J. Manamela, and T. I. Modipa, “HMM-based speech synthesis system incorporated with language identification for low-resourced languages,” in icABCD 2019 - 2nd International Conference on Advances in Big Data, Computing and Data Communication Systems, 2019.
7. N. I. Al-Shathry et al., “MULTI-CLASS SPOKEN LANGUAGE DE- TECTION USING ARTIFICIAL INTELLIGENCE with FRACTAL AL-BIRUNI EARTH RADIUS OPTIMIZATION ALGORITHM,” Frac-tals, vol. 32, no. 9-10, 2024.
8. Y. K. Muthusamy, R. A. Cole, and B. T. Oshika, “THE OGI MULTI- LANGUAGE TELEPHONE SPEECH CORPUS,” in 2nd International Conference on Spoken Language Processing, ICSLP 1992, 1992, pp. 895–898.
9. R. Carlson, B. Granstro¨m, and S. Hunnicutt, “A multi-language text-to- speech module,” in ICASSP, IEEE International Conference on Acous- tics, Speech and Signal Processing - Proceedings, 1982, pp. 1604–1607
10. A. Barve, Y. Ghule, P. Madhani, P. Potdukhe, and K. Pawar, “Multi- language Audio-Visual Content Generation based on Generative Ad- versarial Networks,” in 2023 IEEE World Conference on Applied Intelligence and Computing, AIC 2023, 2023, pp. 33–38.
11. M. Wyawahare, M. Shelke, S. Bhorge, and R. Agrawal, “AI Powered Multilingual Meeting

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

Summarization," in Proceedings of the 14th Inter- national Conference on Cloud Computing, Data Science and Engineer- ing, Confluence 2024, 2024, pp. 86–91.

12. J. R. Tucker and J. F. Cross, "Multi-language presentation of e- learning/distance course content," in Proceedings - 16th Annual General Assembly and Conference of the International Association of Maritime Universities, IAMU AGA 2015, 2020, pp. 367–370.
13. A. Yaganteeswarudu and B. V. Devi, "The multi-language audio compiler with video help," in 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, RTEICT 2018 - Proceedings, 2018, pp. 204–208.
14. E. Oncins, O. Lopes, P. Orero, and J. Serrano, "All together now: A multi-language and multi-system mobile application to make live performing arts accessible," Journal of Specialised Translation, no. 20,
15. pp. 147–164, 2013.
16. D. Konstantinidis et al., "Developing accessibility multimedia services: The case of EasyTV," in ACM International Conference Proceeding Series, 2020, pp. 280–287.
17. R. V. Valentim, I. Drago, M. Mellia, and F. Cerutti, "X-squatter: AI Multilingual Generation of Cross-Language Sound-squatting," ACM Transactions on Privacy and Security, vol. 27, no. 3, 2024.
18. M. I. Saleem et al., "A Novel Machine Learning Based Two-Way Com- munication System for Deaf and Mute," Applied Sciences (Switzerland), vol. 13, no. 1, 2023.
19. N. Klein, T. Chen, H. Tak, R. Casal, and E. Khoury, "Source Tracing of Audio Deepfake Systems," in Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2024, pp. 1100–1104.
20. Y. Wan, R. Zhou, and Y. Yan, "Fast and precise automatic music/speech segmentation," Qinghua Daxue Xuebao/Journal of Tsinghua University, vol. 53, no. 6, pp. 878–882, 2013.
21. A. Bekarystankyzy, A. Razaque, and O. Mamyrbayev, "Integrated end- to-end multilingual method for low-resource agglutinative languages using Cyrillic scripts," Journal of Industrial Information Integration, vol. 43, 2025.
22. Y. Okada, W. Shi, and K. Kaneko, "OpenVSLAM-Based Development Framework for Web-Based VR Tours Using 360VR Videos and Its Extensions," in Lecture Notes on Data Engineering and Communications Technologies, vol. 193, 2024, pp. 31–42.
23. W. Meng and N. Yolwas, "A Study of Speech Recognition for Kazakh Based on Unsupervised Pre-Training," Sensors, vol. 23, no. 2, 2023.