A Computer-Assisted Interpreting System for Multilingual Conferences Based on Automatic Speech Recognition

MS. Preethi P
Assistant Professor of Department, Computer Science and Engineering,
St. Joseph College of Engineering, Chennai-602117, TamilNadu.
Email id – ppreethi@stjoseph.ac.in

Mr. Sakthivel P
Student, Computer Science and Engineering,
St. Joseph College of Engineering, Chennai-602117, TamilNadu.
Email id – sakthinav9655@gmail.com

Mr. Jershin C
Student, Computer Science and Engineering,
St. Joseph College of Engineering, Chennai-602117, TamilNadu.
Email id –jershinchan2003@gmail.com

ABSTRACT - The Computer-Assisted Interpreting System (CAIS) enhances multilingual communication by leveraging cutting-edge AI technologies. In addition to deep learning-based ASR models, the system incorporates speaker diarization to distinguish multiple speakers, preserving conversational flow. It also adapts to various accents and speaking speeds, improving transcription accuracy. Advanced noise filtering ensures reliable speech recognition even in noisy conference environments. CAIS integrates predictive buffering to minimize latency, delivering near-instant translations. Furthermore, it supports domain-specific customization, allowing tailored language models for specialized fields like medical, legal, or technical conferences. With cloud and edge computing capabilities, CAIS provides scalable, real-time multilingual translation for global conference settings.

I. INTRODUCTION

Multilingual conferences require accurate real-time interpretation, but human interpreters are costly, have limited availability, and can experience fatigue, affecting accuracy. Automated solutions like speech-to-text and machine translation face challenges with noise, accents, and latency, disrupting communication flow.This research introduces a Computer-Assisted Interpreting System (CAIS) integrating deep learning-based ASR and NMT for real-time multilingual communication. Using MATLAB-based ASR models, CAIS ensures high-accuracy speech-to-text conversion, advanced noise filtering, and speaker diarization to distinguish speakers. The transcribed text is processed by an optimized NMT model, delivering real-time translations with minimal delay.CAIS is scalable and seamlessly integrates with conference technologies, supporting cloud and edge computing for flexibility. It enables real-time subtitles, speech synthesis, and domain-specific language customization, making it suitable for specialized fields like medicine, law, and technology. Offering a cost-effective,

efficient, and reliable solution, CAIS has the potential to revolutionize multilingual communication, with future enhancements like sign language recognition improving accessibility.

II. BACKGROUND AND MOTIVATION

A. Overview :

The increasing globalization of business, diplomacy, and academia has led to a growing need for efficient multilingual communication. Traditional interpretation services require skilled professionals who may not always be available for every language pair or event. Furthermore, human interpreters can experience fatigue, leading to potential errors and inconsistencies in translation.

With rapid advancements in artificial intelligence, particularly in ASR, machine translation, and speech synthesis, there is an opportunity to create a more scalable and cost-effective solution. ASR systems have achieved significant improvements in accuracy with the advent of deep learning techniques, allowing real-time transcription with high precision. Neural machine translation (NMT) has also evolved, offering contextually accurate translations across multiple languages.

B. Importance of Sensor Detection:

Sensor detection plays a crucial role in improving the accuracy and efficiency of ASR-based interpreting systems. High-quality microphones and environmental sensors help in noise reduction and speaker identification, enhancing speech recognition accuracy. Advanced sensors can detect variations in tone, pitch, and speaker intent, improving the contextual understanding of spoken language. Additionally, integrating biometric sensors ensures personalized interpretation by adapting to individual speaker characteristics, reducing errors in transcription and translation.

C. Motivation for This Research:

By integrating these technologies, a computer-assisted interpreting system can provide real-time, automated interpretation that enhances accessibility and efficiency. Such a system can democratize multilingual communication, making conferences, business meetings, and educational events more inclusive. Additionally, automation reduces operational costs, improves scalability, and ensures consistent interpretation quality, making it an attractive alternative to traditional methods.

Moreover, the motivation behind this research stems from the need to bridge language barriers in global communication. Real-time, AI-driven multilingual interpretation can foster better

collaboration and inclusivity, ensuring that language differences do not hinder meaningful discussions. With continued advancements in ASR, MT, and sensor detection, this research aims to contribute to the development of intelligent, adaptive interpreting systems that cater to diverse linguistic needs.

III. NOVEL APPLICATIONS

The proposed system offers several innovative applications that extend beyond traditional multilingual conferences:

- **Remote Multilingual Collaboration:** Businesses and academic institutions can leverage the system for global virtual meetings, enabling real-time interpretation across various languages.
- **AI-Powered Smart Assistants:** Integration with AI-driven assistants can provide on-the-go language interpretation, enhancing accessibility in international travel and customer service.
- **Augmented Reality (AR) Integration:** AR devices can display real-time translations as subtitles in multilingual conversations, improving user experience in interactive settings.
- **Healthcare and Telemedicine:** The system can facilitate doctor-patient interactions where language barriers exist, improving communication in medical consultations.
- **Educational Support:** Schools and universities can use the system to assist non-native speakers in lectures, improving comprehension and inclusivity.
- **Government and Diplomatic Meetings** – The technology can be deployed in international summits and diplomatic discussions to facilitate seamless communication between delegates speaking different languages.
- **Customer Support and Call Centers** – Businesses can integrate the system into customer service platforms, allowing real-time multilingual conversations between customers and support agents without human interpreters.
- **Legal and Judicial Assistance** – Courts and legal institutions can use the system for real-time language interpretation during trials, ensuring non-native speakers receive accurate translations of proceedings.
- **Tourism and Hospitality Industry** – Hotels, museums, and tourist attractions can deploy AI-powered interpreting devices to assist international travelers with real-time spoken translations.

IV. ROLE AND POTENTIAL

The role of a computer-assisted interpreting system in multilingual conferences is to bridge language barriers and enhance communication efficiency. By integrating advanced ASR, MT, and TTS technologies, the system ensures seamless real-time interpretation without reliance on human interpreters.

The potential of this technology extends beyond conferences, transforming various industries such as healthcare, law, business, and tourism. The system can be adapted for smart cities, automated customer service, and AI-driven content localization. Additionally:

- **Enhanced Accessibility:** Improves access to multilingual communication for individuals with hearing impairments through live subtitles.
- **Workforce Productivity:** Automates multilingual interactions in corporate environments, reducing reliance on human translators.
- **Cross-Border E-Commerce:** Assists international buyers and sellers by providing instant translation of product descriptions and customer interactions.
- **Public Sector and Government Use:** Enables real-time translation in parliamentary sessions and international policy discussions.
- **Tourism and Hospitality:** Facilitates seamless communication for travelers by integrating into mobile apps and digital assistants.

V. FUTURE RESEARCH DIRECTIONS FOR ENHANCED EDUCATION

A. FUTURE RESEARCH DIRECTIONS

1.Advanced Context-Aware Translation: Enhancing machine translation models to better understand cultural and contextual nuances.

2.Real-Time Emotion Recognition: Incorporating AI to detect speaker emotions and adjust translations accordingly.

3.Edge Computing for Low Latency: Reducing processing delays by leveraging edge computing technologies.

4.Enhanced Speech Recognition for Noisy Environments: Improving ASR models to function effectively in environments with background noise.

5.Multimodal Input Integration: Combining text, voice, and gesture recognition for a more comprehensive interpretation system.

6.AI-Powered Summarization: Implementing AI-driven summarization techniques to condense long discussions into concise outputs.

7.Security and Data Privacy Improvements: Developing advanced encryption and anonymization techniques to protect sensitive conversation data.

B. ENHANCED EDUCATION

1.**Real-Time Multilingual Learning** – Enables students to attend lectures in their native language through instant translation.

2.**Personalized Learning Experiences** – AI-driven models adapt to individual student needs, improving comprehension.

3.**Interactive and Inclusive Classrooms** – Supports diverse student populations by removing language barriers.

4.**Improved Language Acquisition** – Helps students learn new languages with real-time speech-to-text and text-to-speech support.

5.**Support for Hearing-Impaired Students** – Provides instant captions and speech synthesis for accessibility.

6. **Speech Analytics for Student Assessment** – Analyzes pronunciation, fluency, and comprehension to enhance learning outcomes.

7. **Integration with Digital Learning Platforms** – Works with e-learning tools like MOOCs, LMS, and video lectures for multilingual education.

8. **Enhanced Teacher-Student Communication** – Facilitates better engagement between teachers and non-native-speaking students.

9.**Collaboration in Global Education Programs** – Enables international student exchange programs by removing language barriers.

10.**Immersive Learning with AR/VR** – Combines CAIS with virtual reality for interactive, multilingual educational experiences.

VI. CONCLUSION:

A computer-assisted interpreting system based on Automatic Speech Recognition (ASR) provides a scalable and cost-effective solution for multilingual conferences. By leveraging ASR to transcribe speech, Machine Translation (MT) to convert it into different languages, and Text-to-Speech (TTS) to deliver spoken output, such a system facilitates real-time interpretation with remarkable efficiency and accuracy.

This approach significantly reduces the dependency on human interpreters, making multilingual communication more accessible and affordable for global events. Although challenges such as accuracy variations due to accents, contextual errors in translation, and latency issues persist, continuous advancements in artificial intelligence and natural language

processing are addressing these limitations. Future improvements in deep learning models, better speech synthesis, and enhanced contextual understanding are expected to further refine automated language interpretation, making it an indispensable tool for international communication

VIII. REFERENCE

1. Nguyen, T. S., Niehues, J., Stüker, S., & Waibel, A. (2020). *Low Latency ASR for Simultaneous Speech Translation*. This paper discusses strategies for achieving low-latency ASR, which is crucial for simultaneous speech translation in multilingual conference settings. Presented at the *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*.
2. Nguyen, T. S., Niehues, J., Stüker, S., & Waibel, A. (2021). *Super-Human Performance in Online Low-latency Recognition of Conversational Speech*. This research demonstrates achieving super-human performance in online, low-latency recognition of conversational speech, which is pertinent to interpreting systems in multilingual conferences. Presented at the *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*.
3. Kannan, A., Datta, A., Sainath, T. N., Weinstein, E., Ramabhadran, B., Wu, Y., Bapna, A., Chen, Z., & Lee, S. (2019). *Large-Scale Multilingual Speech Recognition with a Streaming End-to-End Model*. This paper presents a streaming end-to-end model for large-scale multilingual speech recognition, addressing challenges in handling imbalanced training data across languages. Presented at the *IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*.
4. Nguyen, T. S., Stüker, S., & Waibel, A. (2020). *Low Latency Sequence-to-Sequence ASR for Simultaneous Speech Translation*. This study explores low-latency sequence-to-sequence ASR models tailored for simultaneous speech translation, enhancing the feasibility of real-time multilingual communication. Presented at the *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*.
5. Sperber, M., & Paulik, M. (2020). *Speech Translation and the End-to-End Promise: Taking Stock of Where We Are*. This paper reviews the progress in end-to-end speech translation systems, discussing their potential and challenges in the context of multilingual communication. Presented at the *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*.
6. Zhang, J., Zong, C., & Zeng, H. (2019). *Improving Simultaneous Translation by Incorporating Pseudo-References*. This research introduces methods to enhance simultaneous translation systems by incorporating pseudo-references, contributing to more accurate and efficient multilingual interpreting. Presented at the *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*.
7. Liu, L., Liu, J., & Liu, C. (2018). *Real-Time Speech Recognition and Translation for Multilingual Conferences*. This paper presents a system designed for real-time speech recognition and translation in multilingual conference settings, leveraging ASR technology to facilitate seamless communication. Published in *IEEE Transactions on Audio, Speech, and Language Processing*.