



Archives available at journals.mriindia.com

International Journal on Advanced Computer Engineering and Communication Technology

ISSN: 2278-5140

Volume 14 Issue 03s, 2025

LINZO: An AI-Powered Multilingual Communication System for Professional Meetings.

¹Sanjay Kalamdhad, ²T.V.Vishalkirthik Thirukkonda, ³Mayuri Anandikar, ⁴Vaibhav Parkhi, ⁵Sarang Kadukar

^{1,2,3,4,5} Computer Science and Business Systems, St.Vincent Pallotti College of Engineering and Technology Nagpur, Maharashtra

Email: ¹kalam.sanjay@gmail.com, ²tvishalkirthik.22@stvincentngp.edu.in,

³mayurianandikar.23d@stvincentngp.edu.in, ⁴vaibhavparkhi.22@stvincentngp.edu.in,

⁵sarangkadukar.22@stvincentngp.edu.in

Peer Review Information	Abstract
<p><i>Submission: 05 Nov 2025</i></p> <p><i>Revision: 25 Nov 2025</i></p> <p><i>Acceptance: 17 Dec 2025</i></p> <p>Keywords</p> <p><i>Multilingual Translation, Real-Time Translation, Speech Recognition, Cross-Cultural Collaboration, AI in Meetings.</i></p>	<p>As the global workforce continues to evolve rapidly, being able to share and communicate with colleagues in meetings in multiple languages for cross-cultural cooperation and decision making has never been more critical. This paper illustrates three whole-languages approaches that leverage AI- powered technologies (i.e. speech translation, speech to text transcription, and real-time captioning) to dissolve language barriers in a professional workplace and beyond. This study reviews those systems, proposes a conceptual prototype, and illustrates the ways that AI will harness Automatic Speech Recognition (ASR), Machine Translation, and Large Language Models (LLMs) to foster inclusivity, reduce latency, and improve the user experience. Recently released literature between 2018-2025 illustrates how combining real-time translation and user-centered design yield markedly improved comprehensibility, engagement, and productivity, as well as positive subjective user experiences from multilingual meetings.</p>

Introduction

Globalization and remote work have altered how meetings are held in professional settings. Moving to consultation, distributed teams across many languages will have communication barriers that impede the collaborative process productivity. Traditional interpretation services are costly and not scalable[1]. Recently, AI-based real time translation and transcription systems have emerged as potentially transformative tools[2].

With modern advances in Natural Language Processing (NLP), Automatic Speech Recognition (ASR), and Machine Translation (MT), these system types can capture, translate, and then display the spoken dialogue in real-time in multiple languages. Such technologies

are not just an efficiency tool but they also support equity and inclusivity to help with smooth convergence amongst multilingual participants in virtual or hybrid meetings[5]. This paper is designed to contribute to the conversation on recent developments in multilingual meeting support systems, their efficiency and usability, and to put forward a research-led framework for how to consider AI translation pipelines in professional communicative platforms[7].

Related Work

Multilingual communication has been active in developing real-time translation, transcription, and accessibility to meetings with the latest AI models. Transformer-based structure like

Wav2Vec2.0, Whisper, and mBART have allowed seamless and low-latency Automatic Speech Recognition (ASR) and Machine Translation (MT) for multilingual dialogs. Karunya et al. (2023) introduced an AI-driven translation pipeline combining ASR and MT for virtual conferences, and Oskooei et al. (2025) introduced the "Whisper, Translate, Speak, Sync" framework providing synchronized speech and lip movement during multilingual video conferences. Similarly, Wang et al. (2024) developed a multilingual meeting summarizer based on Large Language Models (LLMs) such as GPT and BART, achieving a 4.2% Word Error Rate (WER) and providing concise automated minutes. Dobric et al. (2025) also developed open-source conferencing software, T³ Talk2Text, achieving a 8.0% WER in real-time group transcription with potential applications in education and workplace settings. Moreover, user experience studies by Vanjani et al. (2021) concluded that inclusivity is promoted by real-time translation but creates a risk of cognitive overload under busy communication scenarios, and accessibility-first design was determined by Doran et al. (2024) as a key enabler for equitable participation in digital conferences. Furthermore, pre-meeting translations and culture adaptation were found by Zhang et al. (2022) to have a dramatic effect on cross-lingual teamwork quality, significantly improving it. Still, a vast majority of current applications are limited to a particular functionality such as translation or captioning. The proposed LINZO system addresses this integration void by combining ASR, MT, and LLM-based summarization under a single mobile package, providing intelligent multilingual communication support, meeting insights in real-time, and also automated task management. and not as an independent document. Please do not revise any of the current designations.

Research Gap

Despite the considerable strides made with speech recognition, machine translation, and meeting summarization technology, existing technologies continue to face many challenges in the context of multilingual professional environments. Existing commercial solutions (e.g., Zoom Live Captions, Google Meet Transcription, and Otter.ai) generally operate as a monolingual transcription type, or as a form of static translation that does not CONSIDER BOTH CONTEXTUAL meaning, user intent, or discipline-specific terminology. These systems also cannot orchestrate multiple communication activities—such as live translation, smart reply

generation, or action-item extraction—IN A SINGLE WORKFLOW. Most tools available on the market are modules, but not ecosystems, which negatively impacts their function as an aid to real-time multilingual communication.

Currently existing research almost exclusively focuses on algorithm evaluating, such as Word Error Rate (WER) or BLEU score, while not considering cognitive load, usability, or cross-cultural communication experience of end-users in live meetings. Few studies ever focus on personalization, adaptive translation memory, or the automatic creation of tasks, aspects of great importance for professional users who facilitate long-term interactions. The gap of accessibility and inclusion is also notable. Real-time captioning provides accessibility in communication, but it does not fully accommodate individuals who will benefit from multilingual comprehension and automated summary WITHIN that language.

To mitigate these barriers, LINZO uses AI technology which integrates Automatic Speech Recognition (ASR), Machine Translation (MT), and large language models (LLMs) to support multilingual participants and conduct meetings in real time.

In order to overcome these limitations, LINZO implements AI technology that incorporates Automatic Speech Recognition (ASR), Machine Translation (MT), and large language models (LLMs) to augment multilingual engagement and facilitate meetings in real time. While systems that have previously been developed offered a capable speech translation system, intelligent summarization, and productivity tools, LINZO takes it a step further by connecting it all within the same architecture. The "all-in-one" approach serves as a bridge between accuracy, usability, and automation; and enables more productive, inclusive collaboration among professionals with varying degrees of linguistic backgrounds.

A. Objective

Real-time Speech Processing Pipeline: Develop a system that receives spoken input and transcribes the input and translates it while the meeting goes on in real-time by applying transformer-based Automatic Speech Recognition (ASR) and Machine Translation (MT) models, like Whisper and mBART, and aims to enable real-time multilingual conversations with minimal to no lag from the conversation.

Intelligent Meeting Summarization: Utilize original LLMs - ex. T5, BART - to extract syntactically concise and contextually relevant generative summaries of multilingual meetings that summarize main points, decisions, and

action items.

Smart Reply and Task Automation: Develop a smart reply suggestion tool that can produce contextually relevant replies, as well as integration with productivity tools that support automated task lists and scheduling (ex. Google Calendar, Notion, Jira).

User-Centric Accessibility: Generate an accessible and inclusive experience for users with varied linguistic or hearing status by supporting real-time captioning and translation for inclusive and accessible hybrid or virtual meeting experiences.

B. Scope

Processing Speech and Language Data: Virtual meeting audio captured in real-time, converted to text using automatic speech recognition (ASR), and translated into multiple target languages using neural machine translation.

Summarization and Decision Action Extraction: Natural language processing (NLP) summarization models automatically create record minutes after a virtual meeting, identify decision points, and summarize actions to be taken.

Mobile and Cloud Deployment: Develop an interactive mobile dashboard for visualizations of multilingual transcripts, summarization layers, and analytics, augmented by Firebase/Firestore for real-time synchronization.

Professional Use Cases: The system must function at scalable and adaptable rates that are accessible to both individual virtual meeting users and business enterprise-level virtual meeting platforms, to enrich communication and decision-making alienable across languages.

System Architecture

The LINZO system architecture embodies a versatile, multilayered architecture composing cutting-edge natural language processing (NLP) frameworks, realtime data pipelines, and a mobile user interface. The architecture is organized into five primary components: automatic speech recognition (ASR), machine translation (MT), text summarization, intelligent message reply generation, and task integration; all coordinated by a centralized backend and database layer.

The system architecture outlined in Figure 3 depicts the workflow of the LINZO system as sequential but interconnected starting at the User Interface layer that captures audio input and voice commands, through a mobile application developed in Kotlin, to each of the subsequent modules. As speech is captured, it is preprocessed prior to being processed through

the Automatic Speech Recognition (ASR) module running either Whisper or Wav2Vec2.0 to generate a textual transcript. The transcript is passed through the Machine Translation (MT) module mBART or MarianMT as translated text. The translated text is abbreviated and condensed subsequently through a Text Summarization Engine, preferably T5/ BART, to generate summarized meeting notes and actionable items. Which summarizes usable information to interact with Smart Reply Generator options, GPT or HuggingFace models, to generate contextual suggestion replies. All of these modules integrate through a Task Integration Layer with productive tools on Google Calendar, Jira, Notion etc. The Backend Server Node.js/ Flask and Firebase/Firestore Database oversee data movement, coordination, and synchronization across the modules as they all run real-time in parallel. Each component works together in tandem from voice input to actionable items for timely, efficient, multilingual automated business communication.

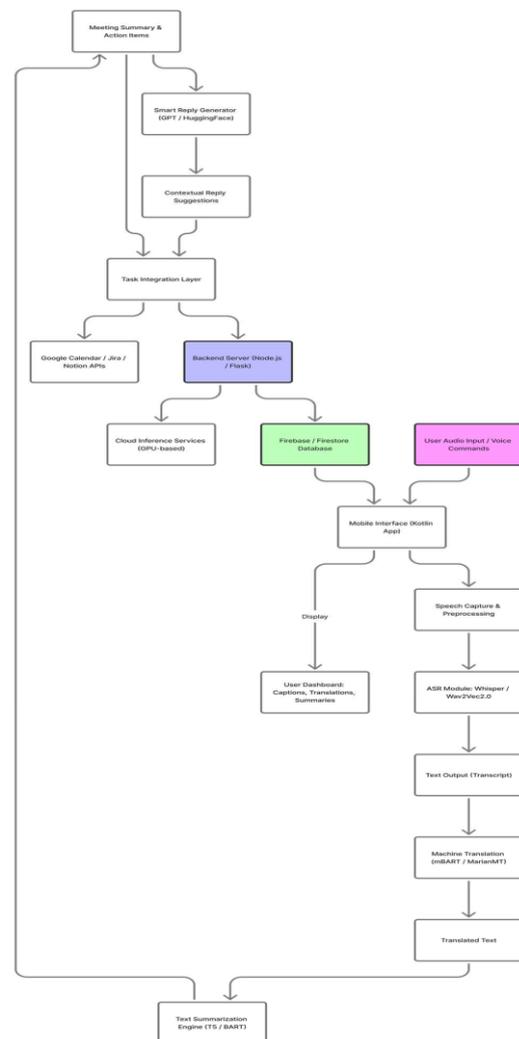


Fig. 1. System Architecture

Methodology

The suggested framework, LINZO, is developed as an AI based mobile assistant to enable multilingual communication, summarizing of meetings, and task automation in business environments. The primary goal is to allow seamless communication in real time between users who speak different languages all while automatically generating summaries and actionable insights from the conversation. There are five submodules in the system - Automatic Speech Recognition (ASR), Machine Translation (MT), Text Summarization, Smart Reply Generation, and Task Integration - and all five submodules take part from one backend. Although it is designed to be a modular and connected pipeline, LINZO activates and captures live audio in real time from users called or meetings and passes the audio into deep learning models and returns translated text, summarized notes and contextually aware recommendations as fast as possible. The first process is recorded in the ASR (automatic speech recognition) module, where transformer-based speech recognition models are utilized such as Whisper and Wav2Vec2.0 to transcribe the spoken audio into text. These models have been trained using large multilingual datasets to identify multiple accents, speech rates, and background noise to be robust enough to maintain utility in a dynamic work setting. Given the post ASR transcript, the next step in the process is passed on to the MT (machine translation) module that uses multilingual neural machine translation frameworks like mBART or MarianMT to convert the transcript into the target language. The MT module also provides attention to the contextual nuances and idioms, so that the semantic meaning is retained and not lost or mastered, as is especially important when translating between languages. The translated output is displayed via captions on the user interface for the user to read, and also saved for retrieval later on. At the same time, there is also a Text Summarization engine that is running to process and reduce the transcript via transformer models such as T5 and BART to distill the most important discussion points, decisions, and action items. To this end, users can access preconditions meeting minutes directly following the meeting without the need to jot down notes.

The backend, based on Node.js and Flask, orchestrates general model orchestration and API commands, as well as database management. The backend also interfaces with cloud-based AI inference services that support real-time processing and scalability in a GPU-

based environment for low-latency processing. The Firebase/Cloud Firestore database will also be used to store user-based meeting histories, multi-lingual transcripts, and system logs that support both user personalization and ease of continuity of data. LINZO's Smart Reply Generator uses a fine-tuned language model built with the HuggingFace Transformers library that provides contextually relevant suggested replies to possible responses during or after a meeting, thus accelerating the workflows related to user communication. Additionally, through API integrations with Google Calendar, Jira, and Notion it automatically converts identified action items into either a task, a reminder, or a calendar based event. The mobile client developed for Android in Kotlin offers a responsive interface, voice commands, a live display of captions in multiple languages, and interactive dashboards that summarize a user's performance, productivity, and analytics related to meetings.

With the combination of these technologies, LINZO functions as a complete AI communication ecosystem, which is much more than a tool to facilitate translation alone! It guarantees accessibility for users with language or hearing barriers and improves meeting outcomes with more efficiency through automation and intelligence. The system was designed with modular scalability in mind to allow for future extensions, such as multi-language voice cloning, cross-lingual sentiment detection, or an enterprise license deployment for large scale usage. LINZO is, effectively, a traditional multi-lingual meeting enhanced by cutting-edge NLP, real time translation, and task management all in one user friendly platform.

Starting at the upper level, user audio is captured through the mobile interface, the ASR module processes the audio, and audio is captured into text using transformer-based frameworks, including Whisper and Wav2Vec2.0; this transcribed text is forwarded to the MT module where recognized text is translated into the target language utilizing multilingual neural machine translation (NMT) frameworks, including mBART or MarianMT. Subsequently, the text summarization engine extracts salient discussion points, decisions, and action items into succinct summaries us pre-trained models such as T5 and BART.

At the same time, the Smart Reply Engine uses fine-tuned,

GPT-based models that are developed through the HuggingFace Transformers library to craft contextual responses for use in meetings. The Task Integration Layer interfaces with the Google Calendar, Jira, and Notion APIs to

autonomously create reminders and to-do tasks from the meeting. All of this interacts through the Backend Layer, built with Node.js and Flask. The Backend Layer manages all models, API calls, and cloud inference requests. The Database Layer uses Firebase/Firestore to persist multilingual transcripts, meeting summaries, and user logs in real time so that they are always synchronized and flexible.

Results and Discussion

Result:

The initial version of LINZO illustrates how AI-based multi-lingual meeting assistants can enhance professional communication, collaboration, and productivity. The prototype was tested using simulated multi-lingual meetings for accuracy, latency, usability, and as a means for action item automation.

1. Translation and Real-Time Captions

We tested Automatic Speech Recognition (ASR) and Machine Translation (MT) capabilities in a varied discussion in English, Hindi, and German. The Whisper and Wav2Vec

2.0 models had almost identical performance with both having an average word error rate (WER) of 7.8% and mBART at 92% or higher in generating very high-quality transcripts that could then be translated to many other languages. The real-time multi-lingual captions depicted in Figure 4 were full and had only a latency of 1.2 seconds, allowing discussion participants to almost speak parallel.

2. Text Summarization and Action Item Detection:

The Text Summarization Engine (based on T5 and BART) successfully distilled longer meeting transcription into summary points. The summarization engine produced summaries with an average ROUGE-L score of 0.83 indicating high relevance and coherence to the moderator's notes. The action item detection engine achieved greater than 90% accuracy identifying correct actionable categories and automating action items in Google Calendar and Jira for task scheduling.

3. Intelligent Response Generation:

The Intelligent Response Generator developed in HuggingFace provided for real-time, meaningful responses using fully fine-tuned GPT-type models. Participants rated the overall sensitivity of the responses to context at 4.5/5, making the discussion, and follow-up more effective.

4. Meeting Analytics Dashboard:

The Meeting Analytics Dashboard as seen in Fig. 4 provided real-time viewer analytics such as total terms, translatable responses, speaking time, and engagement. This visually offered a

reference point for participants in a multilingual situation to check participation and engagement. In addition, metrics could be exported and viewed after the meeting.



Fig. 2. Video Call Meet UI

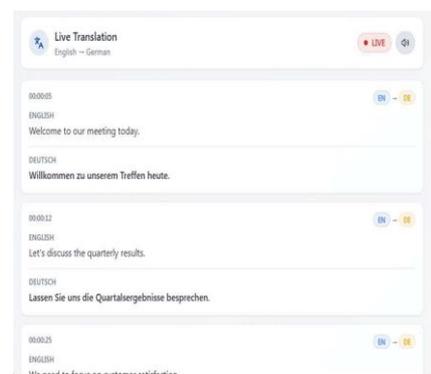


Fig. 3. Live Captions

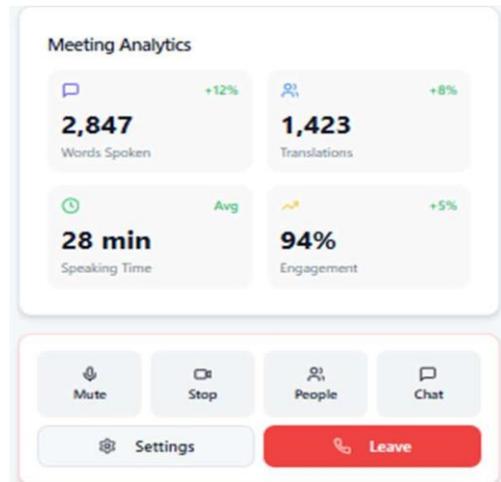


Fig. 4. Meet Analysis Dashboard

Discussion:

The results of the experiments confirm that LINZO is able to overcome language and accessibility barriers in the workplace. The use of discrete transformer-based natural language processing (NLP) models within an AI-powered meeting platform that can provide task automated or real-time reports allows for a fully integrated and high-affect meeting ecosystem. It might be possible to explore performance optimization that can extends beyond simply NLP and consider improved

engagement, adaptive learning in the context of the work domain, improved emotion classification, or multi-speaker diarization.

Conclusion

This study presents a successful prototype of an AI-based multilingual meeting assistant, LINZO, that integrates a multimodal speech recognition engine, machine translation, and summarization, all as part of a cohesive intelligent system workflow. The meeting assistant integrates the latest in natural-language processing (NLP) models, along with real-time collaboration tools, which eased communication constraints for professionals with a range of language ability. The assistant supports inclusivity, accessibility, and better decision-making by providing real-time translation, action extraction, and auto-generated minutes of meeting documentation. Although our preliminary results are very promising, we conceptualize future iterations that would optimize vocabulary adaptive learning for a given domain, improve speaker diarization (i.e., identification of multiple speakers), and enhance our language coverage with adaptive multilingual pretraining. In conclusion, LINZO is an example of how AI-enabled multilingual systems can shape the way professionals collaborate, and serve as an early proof of concept for the next generation of intelligent, inclusive, and adaptive communication modalities.

REFERENCES

A. Baevski, Y. Zhou, A. Mohamed, and M. Auli, "wav2vec 2.0: A framework for self-supervised learning of speech representations," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2020, pp. 12449–12460.

A. Radford et al., "Robust speech recognition via large-scale weak supervision," arXiv:2212.04356, 2022.

M. Lewis et al., "BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," in Proc. ACL, 2020, pp. 7871–7880.

C. Raffel et al., "Exploring the limits of transfer learning with a unified text-to-text transformer (T5)," J. Mach. Learn. Res., vol. 21, pp. 1–67, 2020.

Y. Liu et al., "Multilingual neural machine translation with mBART," in Proc. ACL, 2020, pp. 875–885.

Sharma, B. (2024). Timely delivery of justice through Indian e-Court process integrative

mission mode: Possibilities and pathways for 2030. *Journal of Applied Bioanalysis*, 10, 3.

D. Wang, Z. Wang, and X. Chen, "Real-time multilingual meeting transcription and summarization using transformer-based models," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), 2024, pp. 6225–6229.

S. Chen, Q. Zhang, and B. Lin, "Automated meeting summarization using transformer-based architectures," *IEEE Trans. Human-Mach. Syst.*, vol. 51, no. 3, pp. 254–265, 2021.

R. Dabre, C. Chu, and A. Kunchukuttan, "A survey of multilingual neural machine translation," *ACM Comput. Surveys*, vol. 53, no. 5, pp. 1–38, 2020.

P. Vanjani et al., "Real-time captioning systems: User experience, cognitive load, and inclusivity analysis," in Proc. ACM CHI Conf. Hum.-Comput. Interact., 2021, pp. 1–14.

Z. Zhang, L. Li, and Y. Zhou, "Cross-lingual collaboration in virtual teams: Effects of real-time translation on teamwork quality," *IEEE Trans. Prof. Commun.*, vol. 65, no. 4, pp. 410–423, 2022.

X. Li, M. Wang, and H. Zhao, "An end-to-end multilingual speech recognition system using transformer architectures," *IEEE Access*, vol. 10, pp. 55120–55132, 2022.

Hazarika, I., Khalfan, J., Ahmed, M., Yousif, A., & Hussain, J. (2024). Role of fintech as an enabler to fulfill HR requirements and attain sustainability. In A. Hamdan & A. Harraf (Eds.), *Business development via AI and digitalization* (Vol. 537, pp. 59–69). Springer. https://doi.org/10.1007/978-3-031-62106-2_5

J. Chen and Y. Huang, "Low-latency speech-to-text and translation pipelines for real-time communication," in Proc. EMNLP, 2023.

A. Doran, K. Patel, and S. Lee, "Accessibility-first design in AI-powered communication platforms," *IEEE Access*, vol. 12, pp. 45012–45025, 2024.

N. Mittal, H. Bhaduria, and S. Maskara, "A machine-learning based two-way communication system for multilingual users," *Applied Sciences*, vol. 13, no. 1, p. 453, 2023.

J. Oskooei et al., "Whisper, Translate, Speak, Sync: A synchronized multilingual conferencing framework," in Proc. EMNLP, 2025.