

LIVETRANSMEET: A VIRTUAL MEETING APPLICATION WITH MULTILINGUAL REAL-TIME SPEECH-TO-SPEECH TRANSLATOR AND VIRTUAL BLACKBOARD**Sanjay M. Malode**

Department of Artificial Intelligence and Data Science, K.D.K. College of Engineering, Nagpur, Maharashtra, India,
sanjay.malode@kdkce.edu.in

Aditya Manwatkar

Department of Artificial Intelligence and Data Science, K.D.K. College of Engineering, Nagpur, Maharashtra, India
adityarmanwatkar.aids21f@kdkce.edu.in

Sanskriti Kumar

Department of Artificial Intelligence and Data Science, K.D.K. College of Engineering, Nagpur, Maharashtra, India
sanskritidkumar.aids21f@kdkce.edu.in

Sudhanshu Khodake

Department of Artificial Intelligence and Data Science, K.D.K. College of Engineering, Nagpur, Maharashtra, India
sudhanshurkhodake.aids21f@kdkce.edu.in

Shruti Dhote

Department of Artificial Intelligence and Data Science, K.D.K. College of Engineering, Nagpur, Maharashtra, India
shrutimdhot.aids21f@kdkce.edu.in

Sejal Chiddarwar

Department of Artificial Intelligence and Data Science, K.D.K. College of Engineering, Nagpur, Maharashtra, India
sejalschiddarwar.aids21f@kdkce.edu.in

Abstract

Effective multilingual communication remains a challenge in virtual meetings, especially when participants hail from diverse linguistic backgrounds. To tackle this issue, we have developed LiveTransMeet, a virtual meeting application that integrates dynamic collaboration features with instantaneous speech translation. The application boasts real-time speech translation technology that enables quick translation of spoken English into various target languages, ensuring smooth user interaction. Additionally, it includes an interactive virtual whiteboard for discussing, sharing, and providing feedback on visual content, enhancing both the engagement and effectiveness of meetings. By improving user participation and eliminating language barriers, LiveTransMeet aims to revolutionize online collaboration.

Keywords— Virtual Meeting Application, Real-Time Speech-to-Speech Translation, Multilingual Communication, Interactive Whiteboard, Global Collaboration, SeamlessM4T Model, WebRTC, Natural Language Processing (NLP), Real-Time Communication, LSTM

I. Introduction

Virtual communication tools have become crucial for global business, education, and collaboration in our interconnected world. However, language barriers still pose a significant challenge that often hinders effective communication among diverse, multilingual teams. Existing virtual conferencing platforms, like Zoom, Microsoft Teams, and Google Meet, provide basic communication functionalities but lack integrated solutions to overcome language obstacles and enhance collaborative efforts. LiveTransMeet addresses these challenges by offering a comprehensive virtual meeting platform equipped with advanced features for interactive visual teamwork and real-time

multilingual voice translation. By leveraging cutting-edge technology, LiveTransMeet ensures seamless communication and engagement regardless of users' location or language. The key features of the LiveTransMeet system include:

1. Real-time Speech-to-Speech Translation:

- This feature allows for instantaneous translation of spoken language into the desired language.
- Utilizing Meta AI's pre-trained model, Seamless Streaming, the process combines translation, voice synthesis, and automated speech recognition (ASR) to facilitate natural, real-time interactions across multiple languages.

2. WebRTC Integration for Real-Time

Communication:

- WebRTC ensures low latency and high-quality audio, video, and data transmission directly within web browsers or mobile applications without needing additional software.
- This technology allows all participants to communicate securely and efficiently.

3. Interactive Virtual Blackboards:

- The virtual whiteboards, developed using Python-based frameworks, feature hand gesture recognition supported by programs such as MediaPipe and OpenCV.
 - It is ideal for educational uses and collaborative brainstorming sessions since users can create and work together visually using their natural hand gestures.
 - The addition of an interactive virtual blackboard enhances collaborative experiences by allowing users to share visual content and engage interactively.
- By bringing together these cutting-edge technologies, LiveTransMeet offers a virtual meeting platform that transcends language barriers and encourages visual collaboration. It removes the necessity for separate tools or plugins, ensuring a smooth and engaging user experience. The integration of Seamless Streaming, featuring its strong multilingual translation functions, along with WebRTC's real-time communication infrastructure, establishes LiveTransMeet as a groundbreaking solution for contemporary virtual meetings. Furthermore, the inclusion of a virtual blackboard enhances collaboration, providing intuitive and interactive tools suitable for various applications, including education, team collaboration, and innovative problem-solving.

II. Literature Review

A variety of papers have been examined, each providing distinct insights into the effective implementation of Speech-to-Speech Translation techniques. Below is a summary:

[1] The study titled Seamless-Streaming: Massively Multilingual & Multimodal Machine Translation (Meta AI, 2023) presents an innovative AI model designed for real-time multilingual interactions. It offers support for various types of translations, including speech-to-speech, speech-to-text, text-to-speech, and text-to-text, utilizing a Transformer-based architecture to achieve low latency and high accuracy. This model demonstrates resilience against noise, having

been evaluated with the Fleurs dataset, and it tackles translation bias through the Multilingual Holistic Bias dataset. The assessment of translation quality is conducted using metrics such as Word Error Rate (WER) and chrF. Seamless-Streaming improves worldwide communication with its multilingual and multimodal features, making it particularly suitable for virtual meetings and other real-time scenarios.

[2] On October 25, 2023, a research paper titled "SeamlessM4T: Massively Multilingual & Multimodal Machine Translation" was published. This model is capable of performing automatic speech recognition in as many as 100 languages and supports translations between speech and text in various formats, including speech-to-speech, speech-to-text, text-to-speech, and text-to-text. The development of this model involved using w2v-BERT 2.0 to create self-supervised speech representations from an extensive dataset consisting of 1 million hours of publicly available speech audio.

[3] Ye Jia, Ron J. Weiss, and their team presented a proposal on June 25, 2019, titled "Direct speech-to-speech translation with a sequence-to-sequence model." This research introduces a neural network architecture that allows for translation of speech between different languages without the necessity for an intermediary text representation.

[4] In December 2020, Sireesh Haang Limbu proposed "Direct Speech to Speech Translation Using Machine Learning." This study explores the practicality of translating languages directly in audio format by utilizing a more straightforward Sequence-to-Sequence (STS) encoder-decoder LSTM network, with input provided in the form of spectrograms.

[5] Abhay Kasetwar, Nikita Balani, Deepika Damwani, Alfiya Pandey, Aafreen Sheikh, and Apeksha Khadse presented a proposal in 2022 titled "A WebRTC Based Video Conferencing System with Screen Sharing." In this work, they demonstrated a web-based peer-to-peer real-time communication system utilizing WebRTC technology, HTML5, and a Node.js server address, enabling users to communicate through a channel with high-speed data transfer. The findings indicate that the system is secure, dependable, and fully functional.

[6] Aniket Sandbhor, Prasad Rane, Prathamesh Shirole, and Pawan Phapale introduced "AIR CANVAS" in 2023. This concept revolves

around an air canvas that captures an artist's gestures to draw without the need for a keyboard, mouse, or touchpad contact. The Air Canvas was developed using the Python programming language along with useful libraries like MediaPipe and OpenCV, which were particularly advantageous for tasks related to identification or recognition.

III. Proposed system

A digital conference application named LiveTransMeet was introduced to enhance collaboration and overcome language barriers. It features an interactive virtual whiteboard for gesture-driven drawing, developed using Python and OpenCV, along with real-time multilingual speech-to-speech translation utilizing Meta AI's Seamless-Streaming technique. Additionally, WebRTC facilitates high-quality screen sharing, audio, and video capabilities.

The LiveTransMeet system is a virtual meeting platform that features real-time multilingual speech-to-speech translation along with a virtual blackboard to promote better collaboration.

Here's a summary of its components:

1. Real-time Speech-to-Speech Translation:

- At the heart of the LiveTransMeet system is the Speech Translation Module, designed to eliminate language barriers during live interactions and facilitate smooth multilingual communication. This module offers real-time translations for both speech-to-text and speech-to-speech situations, utilizing Meta AI's Seamless-Streaming AI model. Its pre-trained architecture supports multiple languages, ensuring precise and natural translation outcomes.
- The module begins by capturing audio streams from the microphones of participants, processing them, and converting them into intermediary formats suitable for translation, such as text or spectrograms. The translation process consists of three key stages: Automatic Speech Recognition (ASR), which converts spoken words into text; Language Translation, which employs advanced machine translation models to convert the transcribed text into the desired language; and Text-to-Speech (TTS), which converts the translated text back into spoken language while preserving clarity

and a natural tone.

- Thanks to the low latency optimization, delays in translation are minimized to maintain the flow of conversation. It has been rigorously tested on datasets such as the Fleurs dataset, making it highly effective even in noisy settings, ensuring reliable translations despite challenging audio conditions. The Speech Translation Module is an integral part of the LiveTransMeet platform, recognized for its real-time processing, accuracy, and robustness.

2. Incorporation of WebRTC for Real-Time Communication:

- The foundation of LiveTransMeet's real-time peer-to-peer interaction is the WebRTC-Based Communication Module, which enables high-quality communication by facilitating direct connections among users through the open-source WebRTC (Web Real-Time Communication) framework. This eliminates the reliance on external servers for media routing and supports HD audio and video streaming, dynamically adapting to variable network conditions to ensure a smooth and uninterrupted experience for users.
- Within the video conferencing interface, this module also includes a screen-sharing feature that allows users to display their screens during group discussions, presentations, and demos. Additionally, the module offers participant management functionalities for effective meeting administration, such as creating and overseeing meeting spaces, monitoring attendee engagement, and controlling access and muting participants.
- Alongside video conferencing, there is a text chat feature that permits users to exchange comments, links, and instructions. Furthermore, the module operates in conjunction with the voice translation feature to deliver real-time multilingual text or audio streams, ensuring that translations remain accurately synchronized with the original speech. This synchronization is essential for fostering productive, inclusive, and collaborative virtual meetings while also enhancing communication effectiveness.

3. Collaborative Virtual Blackboards:

- The Virtual Blackboard Module enhances

engagement and productivity in online meetings by enabling real-time collaboration through an interactive, gesture-responsive digital whiteboard. Created using OpenCV and MediaPipe, it incorporates advanced hand gesture recognition that detects users' hand movements in real-time, facilitating simple gesture-based actions such as drawing, selecting colors, and erasing. During meetings, participants can utilize various tools, including pens, erasers, color pickers, and shape-drawing options, to annotate or illustrate on a digital canvas.

- The module allows multiple users to collaborate on the blackboard simultaneously, with real-time updates and sharing of annotations among all

participants fostering dynamic discussions and teamwork. Its seamless integration into the video conferencing experience enables users to switch effortlessly between conversation and collaboration. Interactions on the blackboard are synchronized and broadcast to all meeting attendees via WebRTC.

- The module's cross-platform compatibility guarantees a consistent user experience across devices, functioning flawlessly on desktops, tablets, and smartphones. By empowering participants to annotate shared documents, visually clarify concepts, or brainstorm ideas, it enhances engagement and makes meetings more dynamic, captivating, and productive.

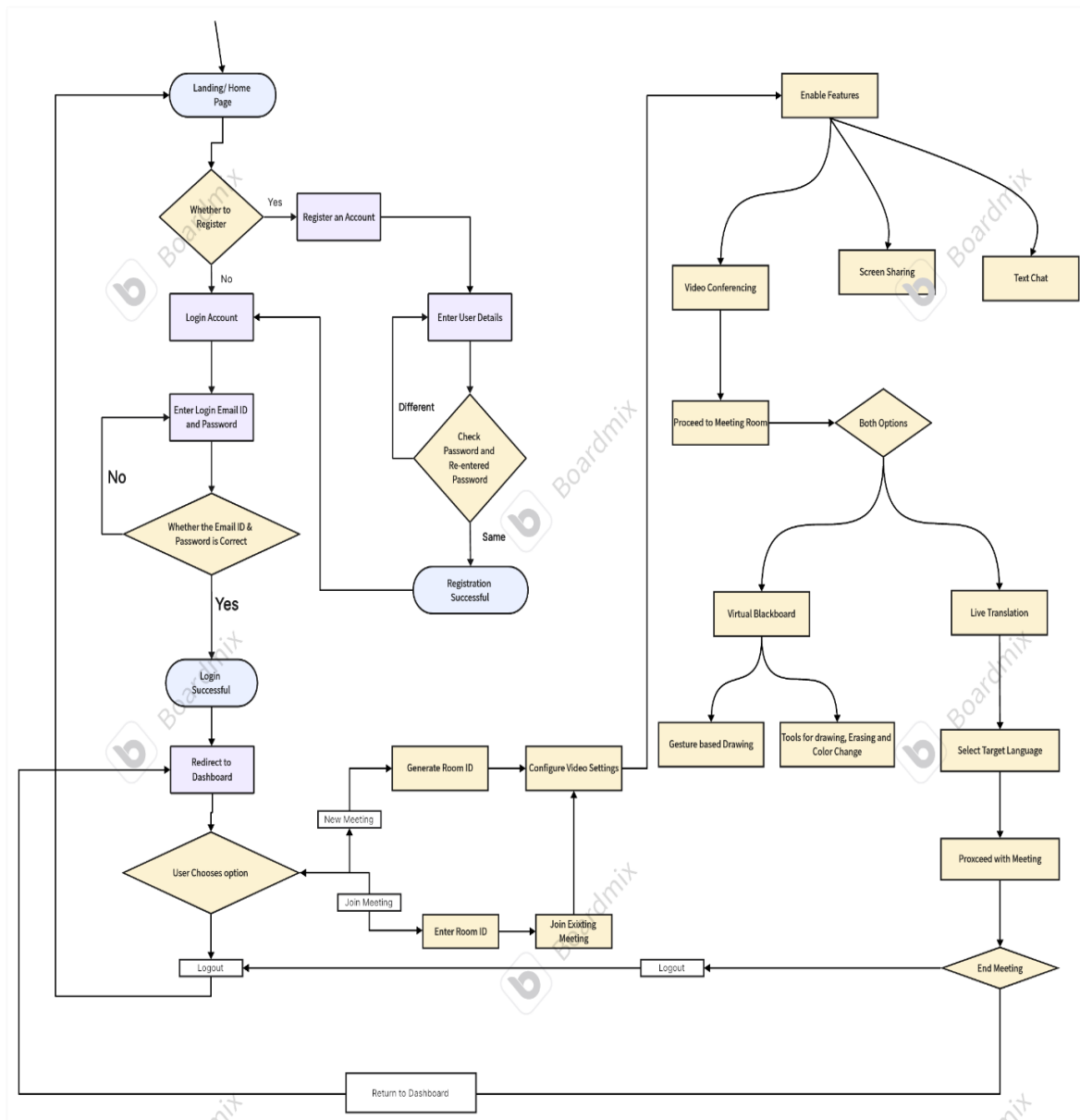


Fig. 3.1 Flowchart for User Interaction and Meeting Flow in LiveTransMeet

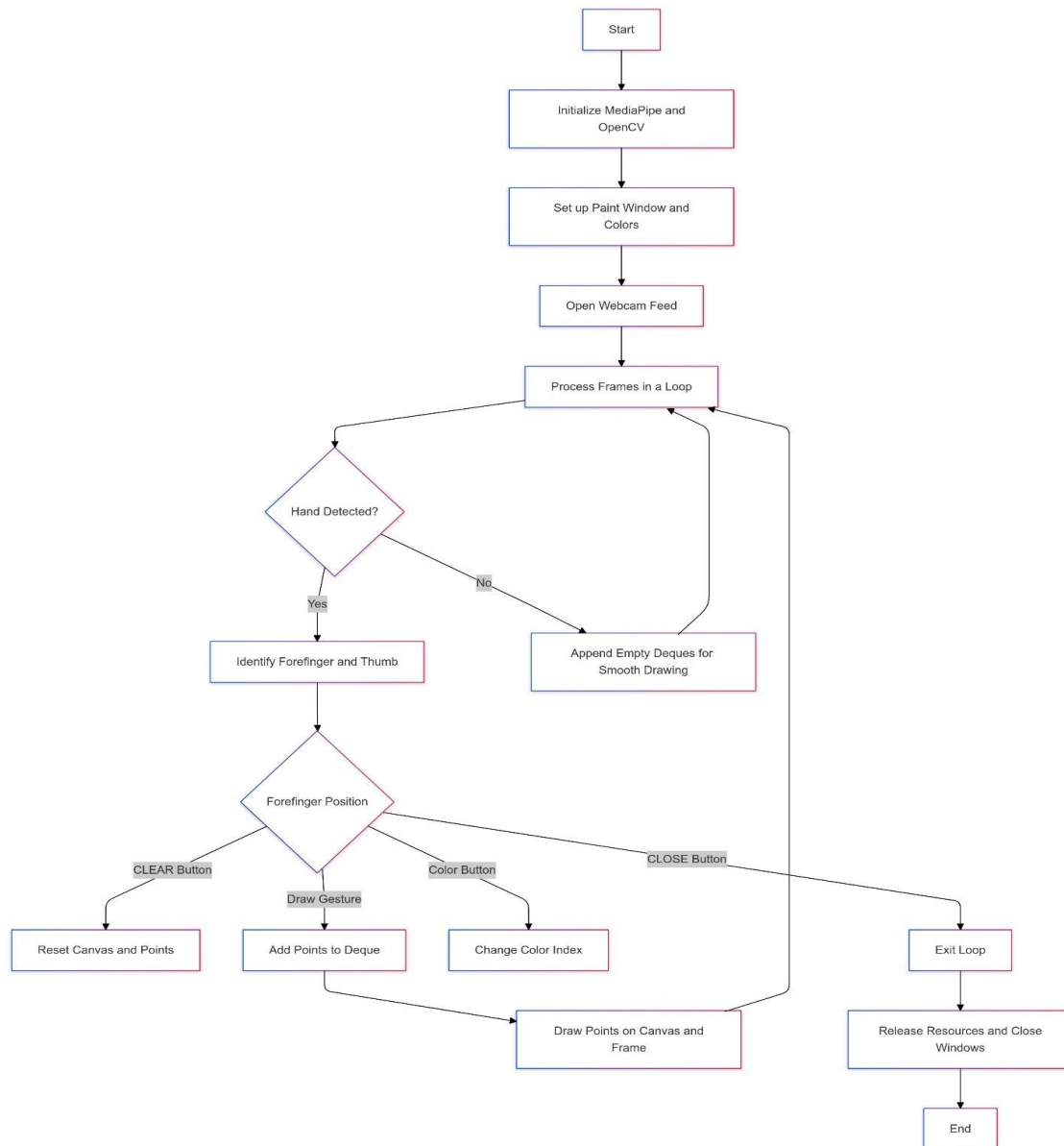


Fig. 3.2 Flowchart for Virtual Blackboard Execution in LiveTransMeet

Alongside meeting the need for cooperative and inclusive online environments, this platform aims to enhance effective communication across language and cultural barriers.

IV. Result

We performed evaluations on LiveTransMeet, a virtual conferencing tool designed to enhance communication among speakers of various languages. The platform incorporates real-time speech-to-speech translation, enabling seamless multilingual interactions. Furthermore, its interactive virtual blackboard enhances collaboration by letting users visually share and discuss concepts. The subsequent results underscore the key features and functionalities of the platform.

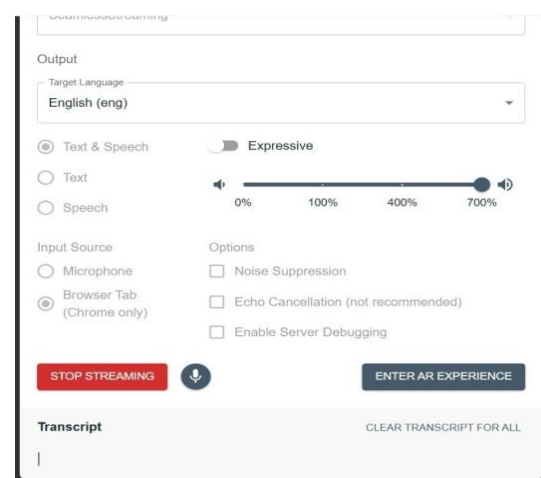


Fig. 4.1 Seamless Streaming Page for Live Translation

The Seamless Streaming Page, which facilitates real-time translation of spoken language, is depicted in this image. Once it captures spoken input and processes it using Meta AI's Seamless Streaming model, the system generates translated speech in the desired target language. The interface allows for effortless multilingual communication during meetings, effectively removing language barriers. To guarantee both accuracy and naturalness in speech delivery, the translation process utilizes Automatic Speech Recognition (ASR), Text-to-Text Translation, and Text-to-Speech (TTS) synthesis. This feature makes LiveTransMeet an essential tool for international conferences, cross-border collaborations, and various online discussions.

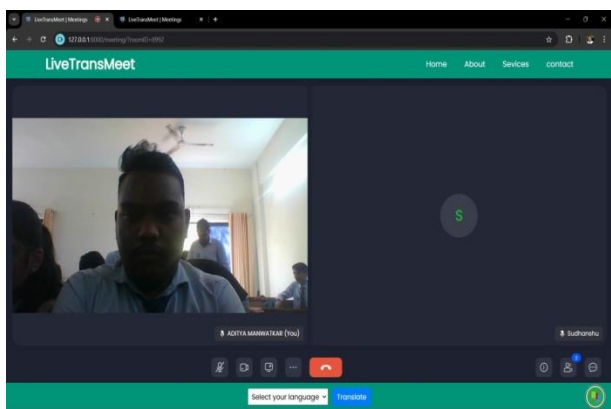


Fig. 4.2 Virtual Meeting Page

The virtual meeting interface, illustrated in fig. 4.2, serves as the main platform for participant communication. By incorporating speech translation to bridge language barriers, it supports real-time audio and video interaction. This platform offers a comprehensive solution for online collaboration, featuring functionalities such as text-based chat, management of participants, and live annotations. Participants can access meetings without the need for software installation due to its intuitive design, ensuring a seamless experience. With its use of WebRTC technology, it is well-suited for both professional and educational environments, providing low latency and high-quality communication.

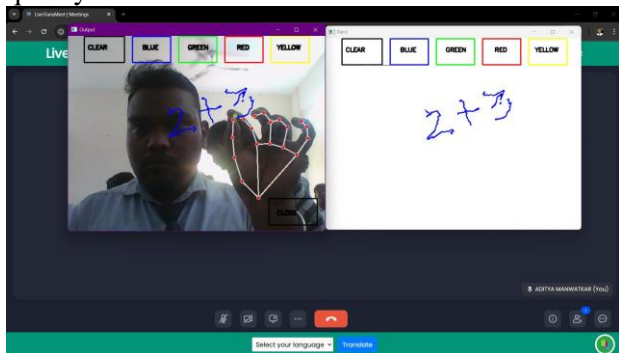


Fig. 4.3 Virtual Blackboard window

By utilizing gesture-based controls, individuals can interact, annotate, and illustrate on the virtual blackboard illustrated in Figure 4.3. This blackboard, developed with MediaPipe and OpenCV for hand gesture recognition, allows users to express their ideas visually with ease. This functionality is particularly beneficial for teachers, students, and professionals seeking an interactive space for brainstorming and discussions. Users can select colors, draw shapes, and erase content through natural hand movements, without needing external tools like styluses. The seamless integration of this blackboard with the meeting interface enhances engagement and makes online meetings more dynamic and productive.

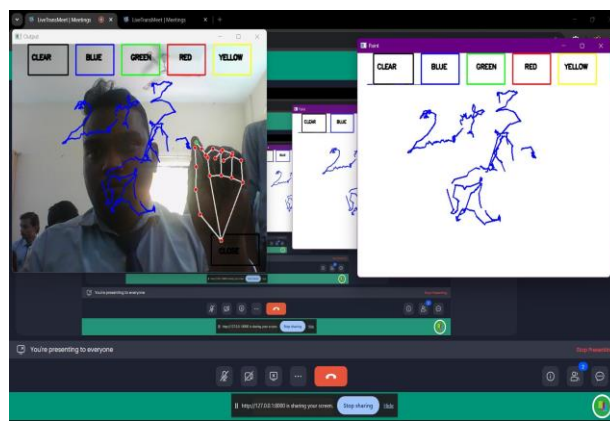


Fig. 4.4 Screen sharing in virtual meeting

Figure 4.4 illustrates the Screen Sharing feature: This functionality enhances presentations and teamwork by allowing users to display their screens in real time. This ensures that participants can access shared information and benefit from interactive features and immediate translation. The feature supports high-resolution screen sharing, maintaining clarity even during detailed lectures, coding demonstrations, and document reviews. Users can smoothly switch between shared displays without disrupting the meeting's flow. Additionally, when used in conjunction with the virtual blackboard, participants can annotate shared content directly, thereby increasing the effectiveness of discussions and knowledge sharing.

V. Conclusion

Tackling the critical issue of language barriers, the "LiveTransMeet" research paper presents an innovative solution in the realm of virtual communication and collaboration. The proposed application aims to enhance cooperation among multilingual teams by integrating real-time speech-to-speech translation with an interactive

virtual whiteboard. By employing advanced technologies such as computer vision, natural language processing, and artificial intelligence (AI), the platform assures accurate translations and dynamic collaborative tools, fostering inclusivity and boosting productivity across various contexts.

One of the key advantages of the project is its focus on robust security measures, a user-friendly interface, and accessibility features; these qualities make it an adaptable solution for international teams, educators, and cross-cultural groups. Alongside comprehensive testing, the iterative process of incorporating user feedback is intended to continually enhance the system and user experience. By enabling effective multilingual interactions and nurturing innovation in both professional and academic settings, this "LiveTransMeet" paper could significantly influence global communication. This platform aspires to create dynamic and diverse environments by prioritizing inclusion, productivity, and flexibility, paving the way for a transformative shift in online collaboration.

VI. Future Scope

The primary objective of the upcoming work plan is to enhance security and privacy. The team aims to implement sophisticated procedures to ensure the confidentiality and integrity of data during virtual meetings and translations since user data is a top priority. This initiative will incorporate measures against unauthorized access, secure authentication methods, and encryption systems. By taking these actions, a trustworthy platform will be created, enabling users to collaborate and communicate without fear of data breaches or misuse.

The next phase involves developing and rigorously testing a multilingual, real-time speech-to-speech translation system. We will leverage advanced artificial intelligence (AI) and natural language processing (NLP) techniques to create a robust translation module. To guarantee precision and reliability, extensive testing in multiple languages is imperative to ensure that the application meets the communication needs of a global audience. This approach will help diminish language barriers, allowing multilingual meetings to proceed more smoothly and efficiently.

The application will be primarily improved based on user feedback. The team plans to gather insights regarding overall user experience, translation accuracy, and ease of use. By making design alterations in response to this feedback,

the application can better meet user preferences and address any concerns. Regular updates informed by actual usage will ensure that the platform evolves with the changing needs of its users, thus enhancing both its functionality and attractiveness.

Finally, the project aims to conduct user experience research to further enhance the functionality and interface. These studies will highlight issues and areas for improvement, leading to a more intuitive and user-friendly design. Moreover, advanced annotation and collaboration features will be integrated into the virtual blackboard function, promoting dynamic user engagement. Collectively, these enhancements will provide "LiveTransMeet" with greater technological capabilities while transforming it into a comprehensive solution for multilingual communication and collaboration across various settings.

References

1. Seamless communication, Barrault, L., Chung, Y. A., Coria Meglioli, M., Dale, D., Dong, N., Duppenhaler, M., ... & Williamson, M., "Seamless:
2. Multilingual Expressive and Streaming Speech Translation.", arXiv:2312.05187v1 8 Dec 2023
3. Meta AI, Seamless communication, Loïc Barrault, Yu-An Chung, et al., "SeamlessM4T: Massively Multilingual & Multimodal Machine Translation", arXiv:2308.11596v3 [cs.CL] 25 Oct 2023
4. Aniket Sandbhor, Prasad Rane, Prathamesh Shirole, Pawan Phapale, "AIR CANVAS", © 2023 IJCRT | Volume 11, Issue 4 April 2023 | ISSN: 2320- 2882
5. Suchitra Deokate, Jagruti Chandivade, Pratiksha Sonawane, Shrija Dange, Preeti Shriwastav, "Real-Time Video Conferencing Application", © 2022 IJCRT | Volume 10, Issue 11 November 2022 | ISSN: 2320-2882
6. Abhay Kasetwar, Nikita Balani, Deepika Damwani, Alfiya Pandey, Aafreen Sheikh, Apeksha Khadse, "A WebRTC Based Video Conferencing System with Screen Sharing", IJRASET ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 10 Issue V May 2022
7. Katherine A. Karl, Joy V. Peluchette, and Navid Aghakhani, "Virtual Work Meetings During the COVID-19 Pandemic: The Good, Bad, and Ugly",
8. Small Group Research 2022, Vol. 53(3) 343 – 365 © The Author(s) 2021
9. Priyanka Padmane, Ayush Pakhale, Sagar

- Agrel, Ankita Patel, Sarvesh Pimparkar, Prajwal Bagde, "Multilingual Speech and Text Recognition and Translation", International Journal of Innovations in Engineering and Science, Vol. 7, No. 8, 2022, PP. 84-88
10. Sireesh Haang Limbu, "Direct Speech to Speech Translation Using Machine Learning", IT 20 088 Examensarbete 30 hp December 2020
 11. Prajwal K R, Rudrabha Mukhopadhyay and et al, "Towards Automatic Face-to-Face Translation", arXiv:2003.00418v1 [cs.CV] 1 Mar 2020
 12. George Suciu, Stefan Stefanescu, Cristian Beceanu, Marian Ceapar, "WebRTC role in real-time communication and video conferencing", DOI: 10.1109/GIOTS49054.2020.9119656
 14. Ye Jia, Ron J. Weiss, Fadi Biadisy, Wolfgang Macherey, Melvin Johnson, Zhifeng Chen, Yonghui Wu, "Direct speech-to-speech translation with a sequence-to-sequence model", Google arXiv:1904.06037v2 [cs.CL] 25 Jun 2019
 15. "A proposed artificial intelligence-based real-time speech to-text to sign language translator for South African official languages for the COVID-19 era and beyond: In pursuit of solutions for the hearing impaired" South African Journal of Communication Disorders ISSN: (Online) 2225-4765, (Print) 0379- 8046
 16. Zinah Nayyef, Sarah Amer and Hussain, "Peer to Peer Multimedia Real-Time Communication System based on WebRTC Technology", International Journal for the History of Engineering Technology, vol. 2, no. 9, pp. 125-130, 2019.
 17. Susanne Wagner (Halle), "Intralingual speech-to-text-conversion in real-time: Challenges and Opportunities" MuTra 2005 – Challenges of Multidimensional Translation: Conference Proceedings © Copyright 2005-2007 by MuTra
 18. https://www.researchgate.net/publication/322686710_Real_Time_Speech_Translation_Google_TV_and_Chat