Grenze International Journal of Engineering and Technology, June Issue

# IntelliCom - Smart and Inclusive Intercom System

Arushi Tiwari[1], Anish Mahadevan[2], KeyurApte[3] and Dr. Rohin Daruwala[4]
[1-4]Department of Electrical Engineering, Veermata Jijabai Technological Institute, Mumbai, Maharashtra
Email: aatiwari_b22@el.vjti.ac.in, amahadevan_b22@el.vjti.ac.in, kvapte_b22@el.vjti.ac.in, rddaruwala@el.vjti.ac.in

***Abstract*—Globally, over 466 million people experience disabling hearing loss, many of whom face significant challenges in traditional workspaces, such as dependence on interpreters, reliance on written communication, and exclusion from spontaneous interactions. IntelliComm is an innovative intercom system designed to enhance accessibility and inclusivity in internal communication, particularly for Deaf individuals. IntelliComm addresses these barriers by integrating sign language detection through video-based gesture recognition, eliminating the need for motion sensors and enabling seamless interaction via a Raspberry Pi 4-powered peer-to-peer video call application. The system supports flexible connectivity through Ethernet, Wi-Fi, and Wi-Fi Direct, allowing offline communication with near-zero latency. Emergency notifications are delivered via vibration motors, ensuring timely and reliable alerts. This paper proposes a scalable, cost-effective solution combining advanced computer vision, efficient application design, and a user-centric approach to bridge communication gaps and foster inclusivity in offices, homes, and similar environments.**

***Index Terms*— Sign Language Detection, Computer Vision, Networking, Intercom, Machine Learning, IoT, Raspberry Pi.**

## I. INTRODUCTION

Disabilities exist only in a world created unequally, where societal structures often fail to accommodate the diverse needs of individuals. For deaf and hard-of-hearing people, communication barriers are a pervasive challenge. Sign language, a natural and effective mode of communication, is not universally understood, limiting interactions to those who are proficient in it. Consequently, Deaf individuals are frequently required to adapt to alternative methods, such as written communication, which is inherently slower, less expressive, and often inadequate in dynamic situations requiring immediacy or nuance.

Traditional solutions, such as Android applications for communication, introduce additional privacy concerns in professional environments where users may be reluctant to share personal phone numbers, or in workspaces where phones may not even be allowed [1]. Intercom systems, however, provide a secure and efficient alternative, enabling seamless communication without compromising individual privacy. Yet, existing intercom systems lack accessibility features tailored for Deaf individuals, leaving a significant gap in inclusivity and usability.

This paper introduces IntelliComm, a peer-to-peer intercom system designed to address these inequities. By leveraging video-based sign language detection, IntelliComm eliminates reliance on interpreters and accelerates communication compared to written text. Its robust offline functionality, privacy-preserving design, and emergency alert features aim to empower Deaf individuals while fostering independence and inclusivity in both personal and professional environments.

*Grenze ID: 01.GIJET.11.2.561_1*

## II. LITERATURE REVIEW

### *A. Intercom System*

Existing solutions have not considered the possibility of an emergency occurring, such as a fire breakout and the deaf person being away from their intercom device [1]. In such a case it is crucial that they are alerted to the dangers immediately. This paper proposes the pairing of each intercom system with a ESP32 powered vibration device that can be used to not only notify the user of meetings and incoming messages, but also alert them during emergencies.

Indian Sign Language (ISL), the sign language that the machine learning model of IntelliCom is trained on has multiple signs which include both static and dynamic signs. Facial features and expression also matter a lot in conveying what a sign means. Since machine learning models cannot accurately mark these parameters in real time, IntelliCom allows 3 person video calls to allow interpreters to join and facilitate much better communication.

Bing-Fei Wu [6] utilizes two separate channels: Ethernet for video and PSTN for voice, ensuring clear communication without bandwidth interference. In a dynamic DHCP environment, the Revolutionary Handshaking Process (RHSP), which employs DTMF signals, handles the exchange of IP addresses between the caller and receiver to set up the connection. However, this method is not as efficient as TCP, which offers reliable data transfer, ensures packets are delivered in order, performs error checking, and adapts to changing network conditions.

Ying-Wen Bai Et al [7] designed a standardized communication protocol using RS-485 that transmits messages in packets, with each packet representing a set of information. To minimize packet loss and errors, the protocol includes an error detection mechanism, ensuring that any erroneous packets are resent.

### *B. Sign Language Recognition*

Various researchers have employed a range of techniques to recognize and interpret different hand gestures or movements, utilizing diverse methodologies and innovative approaches.

ThanekarAadit Et al [2] utilized Convolutional Neural Networks (CNNs) to process colored images by converting them to binary format and resizing them to 64×64 resolution, optimizing computational efficiency while retaining key features. However, the model’s training was limited to 26 letters and a few predefined phrases, and its dependence on a clear background hampers its practicality in real-world scenarios.

Cristian Amaya Et al [3] proposed an automatic sign language recognition system using Principal Component Analysis (PCA) for feature extraction and one-vs-all Support Vector Machine (SVM) for classification. The process begins with hand segmentation through a skin probability model, followed by noise removal using a median filter. Morphological operations enhance the segmented regions, and PCA reduces dimensionality while retaining key features. Finally, the refined features are classified using SVM to accurately recognize hand gestures.

AradhanaKar Et al [4] developed a system that translates ASL videos into simple English sentences. The system extracts frames from each ASL video, stores them in a folder, and processes them using Natural Language Processing (NLP) from an Excel (.xls) file. These processed frames are used to generate grammatically correct simple English sentences via a Sentence Maker module. Finally, the system converts the generated sentences into speech for verbal output.

Shagun Gupta Et al [5] utilize OpenCV for image preprocessing and SURF for feature extraction and key point detection. The system dynamically creates a personalized dataset using input images of the user, enabling more precise gesture recognition. Gestures are matched with this dataset, and a Text-to-Speech (TTS) API converts recognized gestures into audio and text. Implemented in Python, it employs a webcam for gesture capture and offers a user-friendly GUI for customization, allowing real-time communication through intuitive and accurate hand gestures.

## III. HARDWARE

In this intercom system, the core hardware components are designed to enable seamless communication through video and audio, with features like offline sign language detection and speech-to-text conversion. The system leverages Raspberry Pi 4 units, each equipped with essential peripherals and connected via Ethernet for reliable data transfer.

### *A. Raspberry Pi 4B:*

Each unit in the system serves as a standalone communication hub. The Raspberry Pi 4 is chosen for its powerful quad-core ARM Cortex-A72 processor, 8GB of RAM, and connected to each other via Ethernet. This makes it an ideal choice for processing video and audio inputs, handling speech-to-text conversion, and running local sign language detection algorithms.

*B. ESP32-Wroom based vibration device:*

Each intercom system comes paired with a wearable/ desk attachable vibration device that is connected to the intercom system and vibrates to notify the user. The device features an ESP32 Wroom microcontroller dev board which is connected to 2 vibration motors connected to the PWM GPIO output pins of the ESP32. PWM allows for varying intensities of vibration depending on what is to be notified to the user. The aim of this device is to notify deaf people of meeting or alert them of possible emergencies such as a fire breakout. The device is portable and also hosts an OLED screen that displays notifications. The vibration system has a unique MAC-Address and UUID (Unique User Identifier) that allows it to be connected to one intercom system at a time and thus decentralize flow of data, providing better security. The vibration device called VibCom is connected to the Raspberry Pi intercom system using BLE (Bluetooth Low energy), so as to not draw much power and ensure quick alerts.

*C. Camera:*

A USB camera is connected to each Raspberry Pi 4. This camera captures video that can be sent to other Raspberry Pi units for communication. It supports high-definition video capture, making it suitable for video calls.

*D. Microphone:*

A USB microphone array is connected to each unit to capture audio input. This audio input is used for both the speech-to-text feature and voice communication within the intercom system, for ease of communication between people who are not deaf. The Raspberry Pi 4's processing capabilities allow it to perform real-time voice recognition using software like Vosk, which converts spoken words into text for display on the screen.

*E.10-inch touchscreen:*

The device makes use of a 10-inch capacitive touchscreen to allow for proper viewing during video calls and seamless navigation through the custom UI.

*F. Wi-Fi Connectivity:*

The system features full connectivity via Wi-Fi allowing for wireless connectivity and ease of use. Ranges for peer to peer communication can be extended using Wi-Fi extenders or a backend Raspberry Pi can be used to store different IP addresses for connection between devices connected to different wireless networks.

*G. Ethernet Connectivity:*

As a failsafe, to save internet bandwidth, each Raspberry Pi 4 unit can be connected to a local Ethernet network via Ethernet cables. This wired connection ensures stable, high-speed data transfer, which is essential for video calls, real-time audio processing, and the transfer of data between Raspberry Pi devices. The Ethernet interface is used for both video and audio transmission, avoiding the limitations and latency of wireless communication.

*H. Ethernet Switch:*

An Ethernet switch is used to connect multiple Raspberry Pi devices within the intercom system, allowing scalable, reliable communication between them. The switch ensures that each Raspberry Pi is interconnected and can transmit video and audio data to the others without causing network congestion or delays. This setup is crucial for scaling the system, as additional Raspberry Pi units can be added simply by connecting them to the Ethernet switch.

Overall, this intercom system is an integrated, scalable solution that combines video, audio, and sign language recognition using hardware-based components like the Raspberry Pi 4, camera, and microphone. Its use of Ethernet ensures stable, fast communication, while offline capabilities make it a robust option for various communication needs.

## IV. SOFTWARE

The video intercom system integrates real-time sign language detection with custom sign support, and accessibility features tailored for seamless communication. It utilizes Google Mediapipe, Vosk for multilingual speech and sign recognition, and a custom implementation of IP-based video calling using ethernet and Python sockets. Additionally, it features a conference call capability, enabling multiple participants to connect via a centralized or peer-to-peer network. The graphical interface is built on Qt6, providing a robust, and user-friendly experience.

*A.Google Mediapipe for Sign Language Detection*

Google Mediapipe, a framework for building real-time machine learning pipelines, is utilized for recognizing sign language gestures. The system uses the Hands, Holistic, and Pose solutions in Mediapipe to extract key landmarks, enabling efficient and accurate gesture recognition for real-time applications. The detected gestures are converted into text and displayed as subtitles, bridging the communication gap for hearing-impaired users.

*B. IP-Based Communication*

Communication between Raspberry Pi devices is handled over IP using TCP (Transmission Control Protocol). Each Raspberry Pi unit has an unique IP address for ethernet, and video-audio data is transmitted through TCP protocol to ensure reliable flow of data. Video calls are conducted by streaming video data packets over the network, while audio is streamed in parallel to ensure smooth communication.

*C. Vosk for Speech-to-Text and Text-to-Speech*

Vosk, an offline speech recognition toolkit, is integrated into Intellicom for multilingual support in Hindi and English. Vosk's lightweight models allow efficient processing on embedded devices like the Raspberry Pi, providing the following features:
Speech-to-Text: Converts spoken words into text in real-time. For example, a spoken phrase in Hindi or English is transcribed and displayed on the video feed as subtitles, allowing users to follow conversations easily.
The Vosk models for both languages are optimized to work with noisy environments, ensuring robust recognition of the specified language.

*D. Qt6-Based Interface*

The interface developed using Qt6, ensures ease of use and accessibility of which the key features include:
Live Video Feed with Subtitles: Displays ongoing communication with overlaid subtitles from sign and speech recognition.
Call Management: Includes options to start, terminate, and record video calls.
Language Settings: Enables switching between Hindi and English for speech recognition and synthesis.
Qt6's compatibility with embedded platforms ensures smooth performance, even on resource-constrained hardware. The UI features a home screen with an option to register the associated ESP32 vibration device called VibCom. An option to start a meeting, and to send an emergency alert.

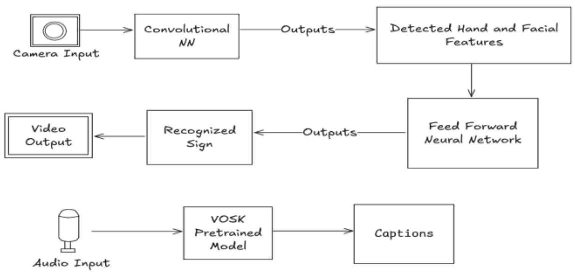


Figure 1. System block diagram

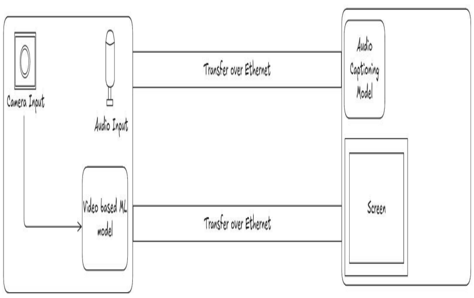


Figure 2. Video Call Block Diagram

*E. System Pipeline*

The software pipeline integrates various modules as follows:
Video Input: The system captures a live stream via the connected camera.
Sign Language Detection: Mediapipe processes video frames to detect predefined and custom signs.
Speech Processing: Vosk converts spoken Hindi or English into text for display.
IP-Based Video Call: The real-time system transmits video, audio, and subtitles over Wi-Fi.
Vibration Device: When the receiver receives the incoming call/ emergency notification, it vibrates at a specified rate according to the message (stronger vibrations for emergency), and notifies the user.

## V. RESULTS

The Smart and Inclusive Intercom System demonstrated effective functionality across its key features, providing a reliable communication solution for individuals with hearing impairments. The vibration motor, triggered in emergency situations, responded within 1 second and proved highly effective in alerting users, even in environments with background noise. System performance was stable over both Wi-Fi and Ethernet connections, with minimal delays in video and audio transmission.
The integration of sign language detection and speech-to-text conversion further enhanced accessibility, with the sign language recognition for letters and numbers achieving 92% accuracy under controlled conditions, and speech-to-text conversion yielding an accuracy rate of 85% in ideal conditions, though performance dropped slightly in noisy environments.
User feedback from deaf and hard-of-hearing individuals, indicated high satisfaction with the system's accessibility features. Participants found the vibration alert useful for emergency notifications, and appreciated the real-time speech-to-text function for live communication. While the system performed well overall, some limitations were noted, including reduced accuracy for complex and highly dynamic signs and occasional transcription errors due to heavy accents or fast speech. Despite these challenges, the system showed significant promise in enhancing both everyday communication and emergency alerts, highlighting its potential as a vital tool for fostering inclusivity.

## VI. CONCLUSION

In conclusion, IntelliCom aims to revolutionize the way we approach communication for the deaf community, especially within traditional workspaces. By incorporating cutting-edge technologies like sign language detection and speech-to-text conversion, we not only enable seamless communication between deaf and non-deaf individuals but also foster an inclusive and supportive environment. This solution not only addresses the communication barriers that often lead to isolation but also ensures that deaf individuals can work more effectively and confidently in their professional settings. Furthermore, with additional features focused on safety, such as emergency vibration alerts tailored to the needs of deaf individuals, IntelliCom ensures that their well-being is prioritized, contributing to a safer and more equitable workspace for all. Ultimately, IntelliCom paves the way for greater inclusivity, helping to build a society where people of all abilities can interact, collaborate, and thrive without limitations.

## FUTURE SCOPE

Indian Sign Language has various dialects and some users may use custom signs to represent different emotions for various words. To encapsulate this a feature can be added to train the model over custom hand gestures that allow the user to communicate with ease. Further, the machine learning model can be improved by taking into consideration facial expressions and having a recurrent neural network to analyse dynamic signs more accurately.

## ACKNOWLEDGEMENT

We would like to express our sincere gratitude to the VJTI (Veermata Jijabai Technological Institute) Alumni Association for supporting our research.
We sincerely express our gratitude to Ms. Isha Barhate, Mst. Aaditya Joil, and Mst. Roshan Adal, our team members for their invaluable contributions to this project. Their efforts in data collection, experimentation, and analysis have been instrumental in drawing the results and capturing the photographs presented in this research paper. Their dedication and support have greatly enriched our work.

## REFERENCES


[1] Gul, B. Zehra, S. Shah, N. Javed and M. I. Saleem, "Two-way Smart Communication System for Deaf & Dumb and Normal People," 2020 International Conference on Information Science and Communication Technology (ICISCT), Karachi, Pakistan, 2020, pp. 1-4

[2] Kar, Aradhana; Chatterjee, PinakiSankar . 2015 International Conference on Computational Intelligence and Networks - An Approach for Minimizing the Time Taken by Video Processing for Translating Sign Language to Simple Sentence in English.

[3] Amaya, C., & Murray, V. (2020). Real-Time Sign Language Recognition. 2020 IEEE XXVII International Conference on Electronics, Electrical Engineering and Computing (INTERCON).

[4] AradhanaKar, PinakiSankar Chatterjee, "An Approach for Minimizing the Time Taken by Video Processing for Translating Sign Language to Simple Sentence in English," in International Conference on Computational Intelligence & Networks, 2015.

[5] Gupta, S., Thakur, R., Maheshwari, V., &Pulgam, N. (2020). Sign Language Converter Using Hand Gestures. 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS).

[6] Bing-Fei Wu, Chao-Jung Chen, Hsin-Yuan Peng, & Cheng-Su Huang. (n.d.). An Innovative Video Phone System in DHCP Network Environment. 2006 Digest of Technical Papers International Conference on Consumer Electronics.

[7] Bai, Y.-W., & Chen, S.-C. (2007). Design and Implementation of Home Intercom and Security Control Systems for Buildings. 2007 IEEE International Symposium on Consumer Electronics.