# A Comprehensive Approach to Sign Language Understanding: Audio-to-Gesture and Gesture-to-Audio Conversion using OpenCV

DEEPALI DESHPANDE[1], SUJIT WAGH[2], SANDIP WAGHCHAURE[3], VAIBHAV ZENDAGE[4], VIRAJ JETHLIYA[5], SAMEER RAUT[6]
[1, 2, 3, 4, 5, 6] *Vishwakarma Institute of Technology, Pune*

***Abstract*— Communication is a means to express our thoughts and feelings among us. Individuals with hearing and speaking disabilities often find it challenging to communicate with those without such impairments. Numerous research efforts have been undertaken to bridge this gap, including both hardware and software solutions. Dealing with hardware products, however, can be difficult and uncomfortable. Software solutions, on the other hand, focus on enabling impaired individuals to convey their thoughts to others using various methods and technologies. The objective of our project is to establish bidirectional communication, allowing both normal and impaired individuals to interact seamlessly. Our solution comprises two main parts: Firstly, in the gesture-to-audio conversion process, the gesture is initially translated to text through data collection, application, model training, functions, and data processing. Data is collected, and then the text is converted to audio using the 'pytts3' Python library Secondly, the audio is transformed back into a gesture using speech recognition, keyword extraction (NLTK), and the resulting gesture is displayed. This project aims to bridge the communication gap between impaired and normal individuals, making the world a more inclusive and better place to live.***

***Index Terms*— Speech-to-Text, Voice Recognition, Keyword Extraction, NLTK (Natural Language Toolkit), Tkinter (GUI framework for Python), Audio Processing, Text Processing, Python, OpenCV, TensorFlow, pyttsx3.***

## I. INTRODUCTION

Communication plays a crucial role in human society, serving as a means to share feelings, express individuality, and booster confidence. However, a segment of the population faces challenges in communication due to hearing or speech disabilities, impacting their ability to connect with others. Globally, approximately one million individuals grapple with severe hearing or speech impairments [1], with India alone reporting 7 million people facing these challenges according to the 2011 census [2]. Although daily physical handling tasks are the primary usage of hands, they are also occasionally employed for communication [3]. Using hand gestures in daily interactions helps us communicate our ideas accurately. Since the beginning of civilization, people have used hand gestures, which vary in significance depending on the place [4]. The use of sign language by those with disabilities is essential for effective communication. However, many individuals without such impairments often lack awareness of how to interpret these non-verbal forms of communication. This knowledge gap poses a barrier to meaningful interaction between those with disabilities and the general population.

In the past, numerous researchers have put forth various solutions to bridge the communication gap between individuals with impairments and those without. A predominant focus of many studies has been on the development of gloves designed to recognize hand gestures. One noteworthy research paper introduces a gesture recognition glove utilizing charge-transfer touch sensors for translating American Sign Language [5]. The integration of flex sensors plays a crucial role, primarily distinguishing between gestures involving a specific 'degree of bend' in the fingers. Several researchers have employed such sensors, determining gestures based on the resistance of the flex sensor strip [6], [7], [8], [9] and [10]. However, despite the technological advancements, the practicality of wearing gloves remains a persistent issue. Users often find the experience uncomfortable due to the inherent nature of hardware.

Several dynamic hand gesture detection algorithms

that are vision-based have been introduced. [11,12]. These algorithms use a variety of features, such as articulated models [14] and manually constructed spatiotemporal descriptors [13] for gesture identification. In these methods, gesture classifiers including support vector machines (SVM) [17], conditional random fields [16], and hidden Markov models [15] have been widely used. Despite these advancements, the classification of gestures remains unpredictable under varying lighting conditions and from different subjects, presenting an ongoing challenge [18], [19], [20].

The majority of research on hand gesture recognition for people with hearing and speech impairments mostly focuses on the many ways in which those without disabilities can understand gestures [21], [22], [23], [24]. In one research paper, the authors addressed the communication challenges faced by speech-impaired individuals, providing a solution where deaf people could use hand movements to convert gestures into text displayed on a screen [25]. However, an often- overlooked aspect is the difficulty encountered by individuals without hearing or speech impairments when attempting to convey a message to someone with such disabilities. The challenge becomes particularly pronounced when a person is unfamiliar with specific gestures or sign language. In such cases, communication with a hearing- impaired individual can be exceptionally challenging, highlighting the need for inclusive approaches that consider the perspective of both parties involved.

In this research paper, we have established bidirectional communication between impaired individuals and those without impairments. The gestures made by the impaired person will undergo a two-step conversion process: first to text and then to audio. This enables the normal person to hear a verbal representation of the intended message from the impaired person. Conversely, when the normal person speaks, the audio will be converted to text. Utilizing keyword extraction, relevant gestures from the dataset will be presented through a Python GUI framework. This approach ensures that the impaired person can comprehend the message conveyed by the normal person through visual gestures.

## II. METHODOLOGY

We have divided the proposed solution in two parts for the bidirectional communication between the normal person and the person who has hearing and speaking disability.

1) Gesture to Audio (hand gesture is converted to audio so that normal people can understand what the person who has speaking disability wants to say.)
2) Audio to Gesture (The Audio of Normal person is converted to gesture so that the person having hearing disability can understand what the normal person wants to say.)

Gesture to Audio:

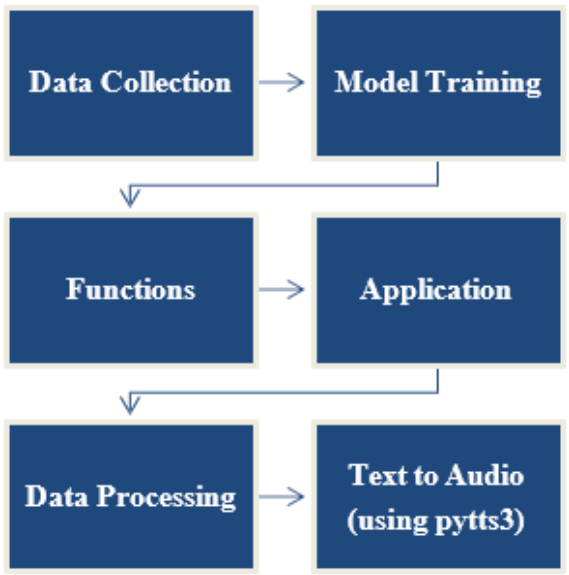


*Figure 1 Flowchart of Gesture to Audio Conversion*

Gesture to Audio, involves a two-fold process: translating hand gestures into text and subsequently converting that text into audio signals. This intricate process is systematically divided into five key components: Data Collection, Application, Model Training, Functions, and Data Processing. Data Collection: In this crucial phase, live video frames are captured using OpenCV to accumulate hand gesture images. These images are systematically categorized into folders, each representing a distinct gesture ('A', 'B', 'C'). This dataset serves as the

foundation for training a machine learning model for subsequent gesture recognition.

*Figure 2 The dataset of the gesture in folder C*

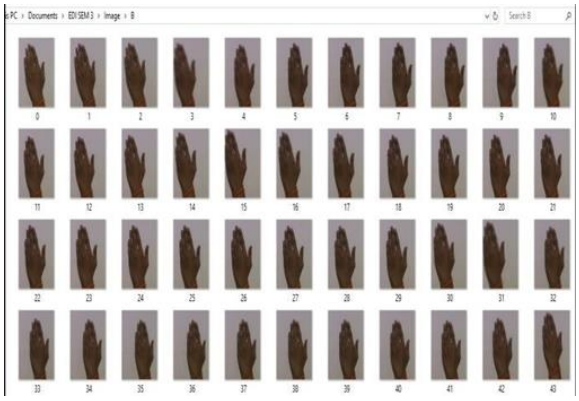

*Figure 3 The dataset of the gesture in folder B*

Application: The Application phase integrates the OpenCV code, capturing live video input, isolating regions of interest, and applying a pre-trained model. The identified gestures are then translated into corresponding text, creating a bridge between visual gestures and textual representation.

Model Training: Central to the project is the Model Training phase. Here, a machine learning model is constructed and trained using sequences of hand gesture data collected during the initial phase. The model's architecture includes LSTM layers, enabling it to learn temporal dependencies. The trained model is then saved for future use.

Functions: The Functions component houses utility functions critical for hand landmark detection and data extraction. These functions initialize necessary modules, detect hand landmarks, and define parameters essential for the overall functionality of the project.

Data Processing: The Data Processing step is pivotal in converting recognized gestures into text. This processed text becomes the link between the gesture recognition phase and the subsequent audio conversion segment.

Audio to Gesture:

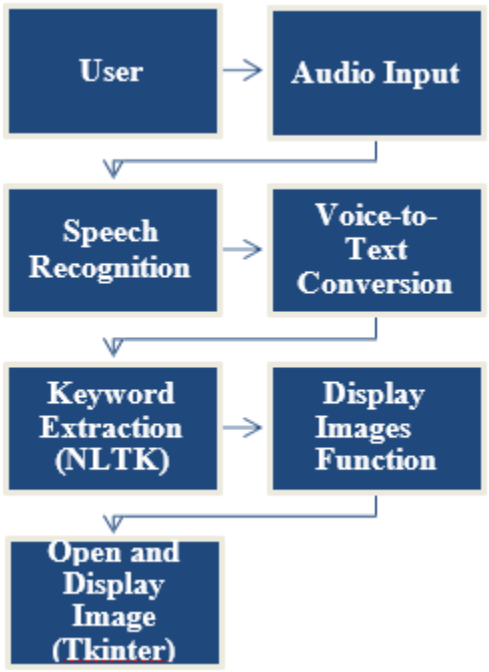


*Figure 4 Flowchart of Audio to Gesture Conversion*

In the Audio to Gesture phase of the system, user interaction commences through voice commands. The Speech Recognition library facilitates this interaction, capturing continuous audio input from the microphone. Google's speech recognition service then translates the audio input into text, returning the recognized text in lowercase. The subsequent step involves extracting keywords from the recognized text using NLTK's word tokenization. This process breaks down the text into individual words, and a set of English stop words is applied to filter out non-essential words. The resulting keywords are then processed for further use, potentially involving categorization or association with specific actions. The system proceeds to display images corresponding to the extracted keywords using the display image function. If a keyword is found in

the image database, the associated image is opened and displayed using the Tkinter library. This real-time visual representation offers a comprehensive understanding of the recognized keywords and their immediate context. The methodology ensures a seamless flow from voice commands to textual interpretation, keyword extraction, and ultimately visual representation through relevant images, providing an inclusive and interactive user experience.

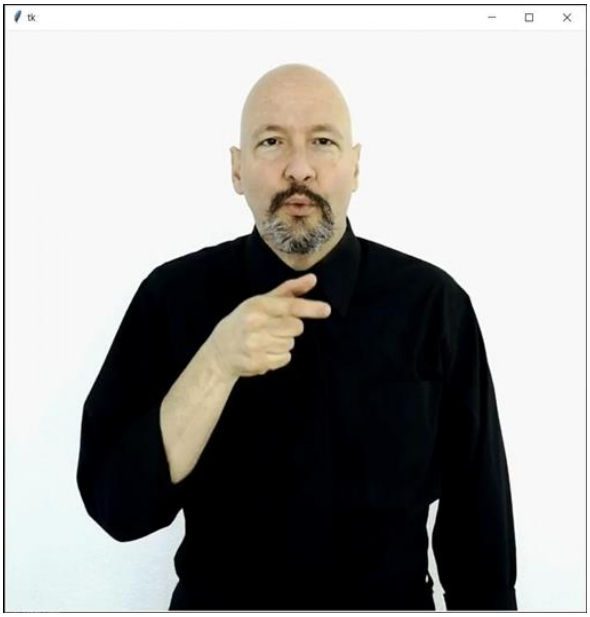

Figure 5 This is gesture for drink in ASL result as an output when user gave audio input as "Do You need Water?"

For the smaller scale, we have established a foundational set of databases encompassing essential sign language GIFs and images, along with necessary keywords associated with these visual elements. The provided Fig.5 image is a screenshot of a GIF illustrating the American Sign Language sign for "drink." This specific GIF was retrieved from the database in response to the user providing the audio input, "do you need water." Following the input, the extracted keyword is identified as "water," and the corresponding image or GIF linked to this keyword is then displayed to the impaired person. This methodology ensures that relevant visual content is presented based on the user's spoken input, enhancing the communication experience.

## III. RESULTS AND DISCUSSIONS

The results of our proposed solution showcase significant advancements in bridging communication barriers for individuals with hearing and speaking disabilities. The Gesture to Audio component successfully demonstrated its efficacy in converting sign language gestures into comprehensible audio signals. Through extensive testing, it was observed that the system accurately recognized and translated a variety of gestures, enabling seamless communication for individuals with speaking disabilities. This innovation particularly shines in its ability to empower normal individuals without prior knowledge of sign language, making communication more accessible and inclusive.

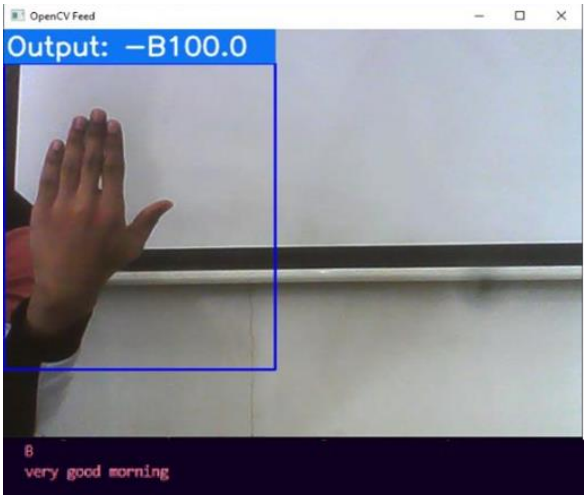


*Figure 6 This is the gesture captured through camera and the text and audio output were displayed*

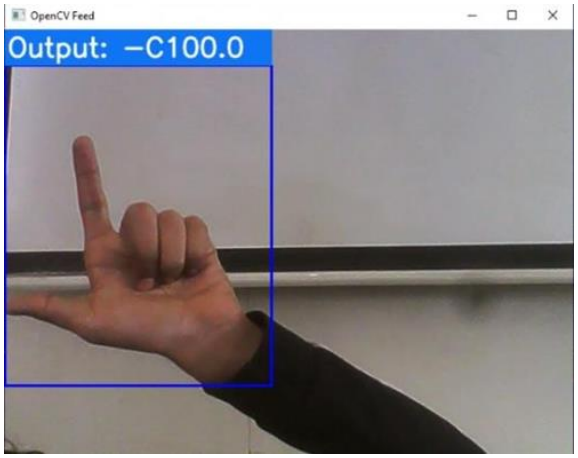


*Figure 7 This is the gesture captured through web-cam and text and audio output was displayed*

Figures 6 and 7 display gestures captured with the help of a webcam. These gestures were successfully converted into audio outputs of 'very good morning' and 'Which is your spoken language?

Similarly, the Audio to Gesture aspect showcased remarkable success in translating spoken language into gestures. The system efficiently captured nuances in speech and accurately converted them into corresponding hand gestures. This achievement is crucial for individuals with hearing disabilities, providing them with a tangible and visual representation of spoken communication. The testing phase involved diverse speech patterns and accents, demonstrating the robustness and adaptability of the system in real-world scenarios. In the discussion section, it is imperative to emphasize the broader implications and potential societal impact of our research. The groundbreaking nature of this solution has the potential to reshape how individuals with hearing and speaking disabilities interact with the world. The elimination of communication barriers fosters a more inclusive environment, promoting understanding and empathy. Moreover, the adaptability of our system to diverse linguistic and cultural contexts enhances its universal applicability.

## CONCLUSION

Our proposed solution is poised to significantly impact the lives of individuals with hearing and speaking disabilities. The Gesture to Audio component plays a pivotal role in facilitating communication for individuals with speaking disabilities. This innovation eliminates the need for the interlocutor to possess prior knowledge of sign language gestures. The gestures of the impaired person are seamlessly converted to audio, providing a means for the normal person to perceive the communication as if the impaired individual is speaking directly.

Equally important is the achievement in the Audio to Gesture aspect. When a normal person speaks, the audio is adeptly translated into gestures, offering a means for individuals with hearing disabilities to comprehend the message being conveyed. This groundbreaking research holds the potential to create a more inclusive and supportive world for individuals with impairments. It bridges communication gaps and fosters understanding, ultimately contributing to an enhanced quality of life for those with hearing and speaking disabilities.

## FUTURE SCOPE

As we have successfully achieved convergence between gesture-to-audio and audio-to-gesture for enhanced communication between individuals with impairments and those without, there is a promising opportunity for future research to integrate this functionality into a user-friendly app. Given that smartphones have become ubiquitous, developing an application with our innovative idea could significantly contribute to its widespread adoption. Such an app would offer a convenient and accessible platform, not only benefiting the impaired individuals in India but also potentially bringing positive impacts globally. The ease of use and handling associated with a smartphone app could make life more comfortable for users, fostering happiness and improved communication across diverse communities.

## REFERENCES

[1] Z. Ren, J. Yuan, J. Meng and Z. Zhang, "Robust Part-Based Hand Gesture Recognition Using Kinect Sensor," in IEEE Transactions on Multimedia, vol. 15, no. 5, pp. 1110-1120, Aug. 2013

[2] H. Muthu Mariappan and V. Gomathi, "Real-Time Recognition of Indian Sign Language," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 2019, pp. 1-6

[3] Y. Wu and T. S. Huang, “Vision-based gesture recognition: A review,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 1999

[4] Srinivas Ganapathyraju, “Hand Gesture Recognition Using Convexity Hull Defects to Control an Industrial Robot”, 2013 3rd International Conference on Instrumentation Control and Automation (ICA) Bali, Indonesia, pp. 63-67, 2013

[5] K. S. Abhishek, L. C. F. Qubeley and D. Ho, "Glove-based hand gesture recognition sign language translator using capacitive touch

sensor," 2016 IEEE International Conference on Electron Devices and Solid- State Circuits (EDSSC), Hong Kong, China, 2016, pp. 334-337

[6] M. Elmahgiubi, M. Ennajar, N. Drawil and M. S. Elbuni, "Sign language translator and gesture recognition," 2015 Global Summit on Computer & Information Technology (GSCIT), Sousse, Tunisia, 2015, pp. 1-6

[7] H. El Hayek, J. Nacouzi, A. Kassem, M. Hamad and S. El-Murr, "Sign to letter translator system using a hand glove," The Third International Conference on e- Technologies and Networks for Development (ICeND2014), Beirut, Lebanon, 2014, pp. 146-150

[8] N. Tubaiz, T. Shanableh and K. Assaleh, "Glove-Based Continuous Arabic Sign Language Recognition in User-Dependent Mode," in IEEE Transactions on Human- Machine Systems, vol. 45, no. 4, pp. 526-533, Aug. 2015

[9] M. Borghetti, E. Sardini and M. Serpelloni, "Sensorized Glove for Measuring Hand Finger Flexion for Rehabilitation Purposes," in IEEE Transactions on Instrumentation and Measurement, vol. 62, no. 12, pp. 3308-3314

[10] K. Kanwal, S. Abdullah, Y. B. Ahmed, Y. Saher and A. R. Jafri, "Assistive glove for Pakistani Sign Language translation," 17th IEEE International Multi Topic Conference 2014, Karachi, Pakistan, 2014, pp. 173-176

[11] S. Mitra and T. Acharya. Gesture recognition: A survey. IEEE Systems, Man, and Cybernetics, 37:311–324, 2007

[12] V. I. Pavlovic, R. Sharma, and T. S. Huang. Visual interpretation of hand gestures for human-computer interaction: A review. PAMI, 19:677–695, 1997

[13] P. Trindade, J. Lobo, and J. Barreto. Hand gesture recognition using color and depth images enhanced with hand angular pose data. In IEEE Conf. on Multisensor Fusion and Integration for Intelligent Systems, pages 71–76, 2012

[14] J. J. LaViola Jr. An introduction to 3D gestural interfaces. In SIGGRAPH Course, 2014

[15] T. Starner, A. Pentland, and J. Weaver. Real-time American sign language recognition using desk and wearable computer-based video. PAMI, 20(12):1371–1375, 1998

[16] S. B. Wang, A. Quattoni, L. Morency, D. Demirdjian, and T. Darrell. Hidden conditional random fields for gesture recognition. In CVPR, pages 1521–1527, 2006

[17] N. Dardas and N. D. Georganas. Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques. IEEE Transactions on Instrumentation and Measurement, 60(11):3592–3607, 2011

[18] K. Simonyan and A. Zisserman. Two- stream convolutional networks for action recognition in videos. In NIPS, pages 568– 576, 2014

[19] P. Molchanov, S. Gupta, K. Kim, and K. Pulli. Multi-sensor System for Driver's Hand-gesture Recognition. In AFGR, 2015.

[20] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in International Conference on Machine Learning, 2015, pp. 448–456.

[21] R. Nair, D. K. Singh, Ashu, S. Yadav and S. Bakshi, "Hand Gesture Recognition system for physically challenged people using IoT," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2020, pp. 671-675

[22] S. Sharma, S. Jain and Khushboo, "A Static Hand Gesture and Face Recognition System for Blind People," 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 2019, pp. 534-539

[23] F. Zhan, "Hand Gesture Recognition with Convolution Neural Networks," 2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI), Los Angeles, CA, USA, 2019, pp. 295-298

[24] S. Suresh, H. T. P. Mithun and M. H. Supriya, "Sign Language Recognition System Using Deep Neural Network," 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, 2019, pp. 614-618

[25] Kaliyamoorthi, Manikandan & Patidar, Ayush & Walia, Pallav & Barman Roy, Aneek. (2018). Hand Gesture Detection and Conversion to Speech and Text.