Real-Time Indian Sign Language(ISL)Recognition And Multilingual Translation System Using Deep Learning And Natural Language Processing

Nivetha S M[1], Obuli Dharani Dharan O[2], Akash S [3], Arunprasath S[4], Mouleeshwaran G[5]
[1]Assist prof, Dept of Information Technology,
[2, 3, 4, 5] Dept of Information Technology,
[1, 2, 3, 4, 5] Dhirajlal Gandhi College of Technology,Salem,TamilNadu.

Abstract- *Artificial Intelligence (AI) has become a powerful tool in developing assistive technologies that improve accessibility for individuals with disabilities. Among these, hearing and speech-impaired individuals face significant challenges in communication due to the lack of widespread understanding of Indian Sign Language (ISL). ISL is a visual language that relies on gestures, facial expressions, and body movements. However, most people are not familiar with it, leading to communication barriers in everyday life. Existing systems mainly focus on recognizing individual alphabets or static gestures, which limits their ability to provide real-time, meaningful communication.*

This paper proposes an AI-based two-way communication system designed to bridge the gap between hearing-impaired individuals and others. The system converts ISL gestures into text and speech while also translating spoken language into text, enabling bidirectional communication. The proposed approach integrates Computer Vision, Machine Learning, and Natural Language Processing techniques. MediaPipe is used for capturing real-time hand landmarks, while Long Short-Term Memory (LSTM) networks are employed for recognizing dynamic gesture sequences. Recognized gestures are mapped into gloss representations, which are further processed into meaningful sentences using NLP models such as BART. Additionally, multilingual translation is achieved using IndicTrans2, and speech output is generated using Indic Text-to-Speech systems.

The system is designed to be efficient, cost-effective, and user-friendly, making it suitable for real-world applications. The results demonstrate improved accuracy, real-time performance, and better contextual understanding compared to existing methods. This research contributes to enhancing accessibility and promoting inclusivity by enabling effective communication between hearing and non-hearing communities.

Keywords: Artificial Intelligence, Indian Sign Language, Machine Learning, Computer Vision, NLP, Assistive Technology

I. INTRODUCTION

Communication is essential for human interaction, enabling individuals to share ideas, thoughts, and emotions. However, for hearing and speech-impaired individuals, communication becomes challenging due to the lack of understanding of sign language among the general population. Indian Sign Language (ISL) is widely used by deaf individuals in India, yet it is not commonly understood by most people. This creates a communication gap that affects education, employment, and social interaction.

Traditional communication methods such as interpreters or written text are not always accessible or efficient. Moreover, existing technological solutions are limited in their ability to recognize continuous gestures and generate meaningful sentences. Most systems focus only on alphabet recognition, which is insufficient for real-world communication.

Recent advancements in Artificial Intelligence, particularly in Computer Vision and Natural Language Processing, have opened new possibilities for developing intelligent communication systems. These technologies enable machines to understand visual inputs and convert them into meaningful outputs. This research focuses on leveraging these advancements to create a real-time, bidirectional communication system that can translate ISL gestures into text and speech, as well as convert speech into text.

II. LITERATURE REVIEW

Sign language recognition has gained significant attention in recent years. Various approaches have been proposed using Machine Learning and Deep Learning techniques.

Early systems used traditional image processing methods to detect hand gestures, but these approaches lacked accuracy and robustness. With the introduction of

Convolutional Neural Networks (CNNs), gesture recognition improved significantly. However, CNN-based models are more suitable for static gestures and struggle with dynamic sequences.

To address this limitation, researchers introduced Long Short-Term Memory (LSTM) networks, which are capable of handling sequential data. LSTM-based systems have shown improved performance in recognizing dynamic gestures. Integration with MediaPipe further enhances accuracy by extracting precise hand and body landmarks.

Recent research has explored Transformer-based models and hybrid approaches combining CNN, LSTM, and attention mechanisms. These models achieve higher accuracy and better contextual understanding. However, challenges such as limited datasets, lack of standardization, and real-time processing constraints still exist.

The proposed system addresses these challenges by combining efficient gesture recognition with NLP-based sentence generation and multilingual support.

III. EXISTING METHODS

Existing systems such as SignAll and Hand Talk provide partial solutions but lack support for Indian Sign Language and real-time bidirectional communication. Our proposed system addresses these limitations by integrating Computer Vision, NLP, and multilingual capabilities.

A. Indian Sign Language Research and Training Centre

The Indian Sign Language Research and Training Centre plays a significant role in promoting and standardizing Indian Sign Language (ISL) across the country. It has developed an official ISL dictionary that helps in learning and understanding standardized signs, along with providing datasets and educational resources for researchers and learners.

The organization primarily focuses on the development, training, and dissemination of ISL to improve accessibility for hearing-impaired individuals. However, despite its valuable contributions, ISLRTC does not offer a real-time AI-based translation system, and its resources are mainly limited to learning and reference purposes rather than enabling dynamic, real-time communication.

B. Google Live Transcribe

Google Live Transcribe is an accessibility application that converts spoken language into real-time text using automatic speech recognition. It supports multiple languages and helps hearing-impaired individuals follow conversations easily. The app is widely used due to its simplicity and effectiveness. However, it does not support sign language recognition, as it only processes audio input, limiting its ability to provide complete communication for users who rely on sign language.

C. Hand Talk

Hand Talk is an accessibility application that converts text and audio into sign language using a 3D animated avatar, making communication easier for hearing-impaired users. It is widely used in Brazil and supports Brazilian Sign Language (Libras). The app enhances accessibility in digital platforms, but it does not support Indian Sign Language (ISL) and offers limited bidirectional communication, as it mainly focuses on one-way translation.

IV. PROPOSED METHODOLOGY:

Introduction

The proposed methodology presents a comprehensive approach for developing a real-time, AI-based two-way communication system for Indian Sign Language (ISL). The system integrates Computer Vision, Machine Learning, and Natural Language Processing techniques to enable seamless interaction between hearing-impaired and non-hearing individuals. It is designed to process both visual (gesture) and audio (speech) inputs and generate meaningful outputs in the form of text and speech.

System Architecture and Workflow

The system follows a modular architecture consisting of multiple interconnected components. Initially, gesture input is captured using a webcam, and speech input is recorded via a microphone. The visual data is processed using MediaPipe to extract hand landmarks, which are then passed to an LSTM model for dynamic gesture recognition. The recognized gestures are converted into gloss sequences and further processed using NLP models to generate meaningful sentences. Simultaneously, speech input is converted into text using speech recognition techniques. The final output is delivered as text and speech, enabling bidirectional communication.

Implementation and Practical Considerations

The system is implemented using Python with libraries such as OpenCV for image processing, MediaPipe for hand tracking, and Tensor Flow/PyTorch for model development.

Real-time performance is achieved by optimizing frame processing and reducing computational overhead. The system is designed to work efficiently on standard devices without requiring expensive hardware.

Practical considerations include handling varying lighting conditions, background noise, and different hand orientations. User interface design is kept simple and intuitive to ensure ease of use for all users.

Mathematical Models and Formulation

The proposed system models sign language recognition as a sequential pattern classification problem, where gestures are represented as time-series data of hand landmark coordinates. Each gesture sequence is denoted as a set of feature vectors extracted over time, capturing both spatial and temporal variations. A recurrent neural network, specifically a Long Short-Term Memory (LSTM) model, is employed to learn temporal dependencies and context within the gesture sequence. The model processes input features through hidden states and applies nonlinear transformations to capture complex motion patterns. The final classification is obtained using a probabilistic function, typically softmax, which assigns the most likely gesture label. Additionally, Natural Language Processing (NLP) models transform recognized gesture sequences into grammatically correct sentences, ensuring semantic coherence. This mathematical formulation enables accurate recognition and meaningful interpretation of continuous sign language gestures in real time.

Algorithm

The proposed algorithm begins by capturing real-time video input through a webcam and extracting frames for processing. Hand landmarks are detected using MediaPipe and converted into normalized feature vectors. These sequential features are passed into an Long Short-Term Memory (LSTM) model to recognize dynamic gestures. The predicted gesture labels are mapped into gloss sequences and refined into meaningful sentences using NLP techniques. The system then generates text and speech output, while also processing speech input to convert it into text for enabling bidirectional communication.

System Architecture and Workflow

The proposed system operates in a continuous pipeline where gesture and speech inputs are processed simultaneously. The integration of Computer Vision and NLP ensures accurate recognition and meaningful output generation. The workflow is optimized for real-time communication, making the system practical for real-world applications.

Implementation and Practical Considerations

The system emphasizes efficiency, scalability, and usability. By leveraging lightweight models and optimized processing techniques, it ensures smooth performance on standard devices. Challenges such as lighting variations and gesture complexity are addressed through preprocessing and robust model training.

Mathematical Models and Formulation

The proposed system models sign language recognition as a sequential pattern classification problem, where gestures are represented as time-series data of hand landmark coordinates. Each gesture sequence is denoted as a set of feature vectors extracted over time, capturing both spatial and temporal variations. A recurrent neural network, specifically a Long Short-Term Memory (LSTM) model, is employed to learn temporal dependencies and context within the gesture sequence. The model processes input features through hidden states and applies nonlinear transformations to capture complex motion patterns. The final classification is obtained using a probabilistic function, typically softmax, which assigns the most likely gesture label. Additionally, Natural Language Processing (NLP) models transform recognized gesture sequences into grammatically correct sentences, ensuring semantic coherence. This mathematical formulation enables accurate recognition and meaningful interpretation of continuous sign language gestures in real time.

Algorithm

The proposed algorithm begins by capturing real-time video input through a webcam and extracting frames for processing. Hand landmarks are detected using MediaPipe and converted into normalized feature vectors. These sequential features are passed into an Long Short-Term Memory (LSTM) model to recognize dynamic gestures. The predicted gesture labels are mapped into gloss sequences and refined into meaningful sentences using NLP techniques. The system then generates text and speech output, while also processing speech

input to convert it into text for enabling bidirectional communication.

Overview

The proposed system follows a structured pipeline to transform input data into meaningful and context- aware output. It integrates multiple stages, including input processing, sentiment analysis, contextual understanding, and optimized output generation, ensuring accurate and effective communication.

1. Input Processing
2. Sentiment Detection
3. Contextual Encoding
4. Rewrite Generation
5. Reinforcement Optimization
6. Output Generation

System Architecture

The system architecture is designed as a modular and scalable framework that integrates Computer Vision, Machine Learning, and Natural Language Processing to enable real-time bidirectional communication. It consists of input modules for capturing gestures through a webcam and speech through a microphone. The visual input is processed using MediaPipe to extract hand landmarks, which are then passed to an Long Short- Term Memory (LSTM) model for dynamic gesture recognition. The recognized gestures are converted into gloss representations and further refined into meaningful sentences using NLP models. In parallel, the speech input is processed using speech recognition techniques to generate text output. The final system produces both text and speech outputs, ensuring seamless two-way communication. The architecture is optimized for real-time performance, flexibility, and ease of deployment in practical environment

MATHEMATICAL MODEL

The gesture recognition process is modeled as a sequence classification problem. Let the input sequence of hand landmarks be represented as

$$X = \{x_1, x_2, x_3, \dots, x_t\}$$

where each $x_t$represents the feature vector of hand landmarks at time step t. The LSTM model computes hidden states using:

$$ht = f(W \cdot xt + U \cdot ht - 1 + b)$$

where W, U, and $b$are model parameters. The final output is obtained using a softmax function:

$$y = \text{softmax}(h_t)$$

which predicts the probability distribution over gesture classes.

V. IMPLEMENTATION

The proposed system is implemented using Python by integrating multiple libraries and frameworks to achieve real- time performance. The gesture recognition module utilizes OpenCV for video capture and preprocessing, while MediaPipe is employed to extract precise hand landmarks from each video frame. These landmarks are normalized and fed into a deep learning model based on Long Short-Term Memory (LSTM), which is trained to recognize dynamic gesture sequences in Indian Sign Language. The model is developed using TensorFlow or PyTorch, ensuring efficient training and inference.

For sentence generation, recognized gestures are mapped into gloss representations and processed using Natural Language Processing models to form grammatically correct sentences. The system also integrates speech recognition APIs to convert spoken input into text, enabling reverse communication. Additionally, text-to-speech functionality is incorporated to generate audio output from the predicted text.

To ensure real-time usability, optimization techniques such as frame rate control, lightweight model design, and efficient memory handling are applied. The system is designed to run on standard computing devices without requiring specialized hardware, making it cost-effective and practical. A simple and user-friendly interface is developed to allow easy interaction for both hearing-impaired and non-hearing users.

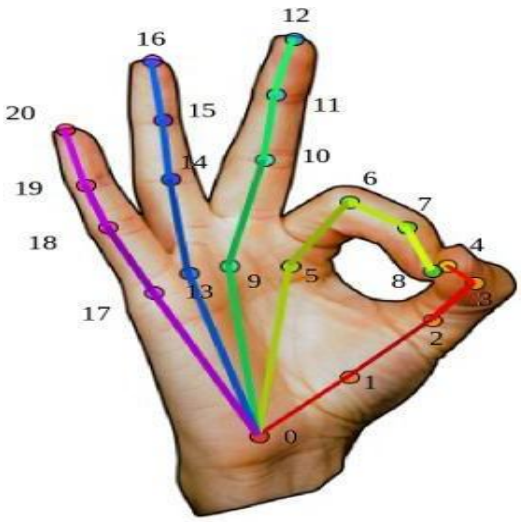

Figure 1: ISL Diagram

VI. RESULTS AND ANALYSIS

A. Performance Improvement

The proposed system shows significant improvement over traditional sign language recognition approaches. By integrating MediaPipe for precise feature extraction and Long Short-Term Memory (LSTM) for sequential modeling, the system achieves higher accuracy in recognizing dynamic gestures compared to static CNN-based models. Real-time performance is enhanced through optimized processing and lightweight model design, reducing latency and improving response speed.

Additionally, the use of NLP techniques enables context-aware sentence generation, which improves the overall quality and usability of the output.

B. Sample Output

The system produces outputs in both text and speech formats. For example, a gesture sequence representing “How are you” is accurately recognized and displayed as text, followed by corresponding audio output. Similarly, spoken input such as “I need help” is converted into text for the hearing-impaired user. The outputs are clear, contextually accurate, and generated in real time, demonstrating effective bidirectional communication.

C. Key Observations

The experimental results highlight several important observations. The system performs well under normal lighting and controlled environments, maintaining high accuracy and responsiveness. The combination of Computer Vision and NLP significantly improves sentence formation compared to traditional gesture recognition systems. However, performance may slightly degrade under poor lighting conditions, complex backgrounds, or very fast hand movements. Despite these limitations, the system proves to be reliable, scalable, and suitable for real-world communication, offering a substantial improvement over existing solutions.

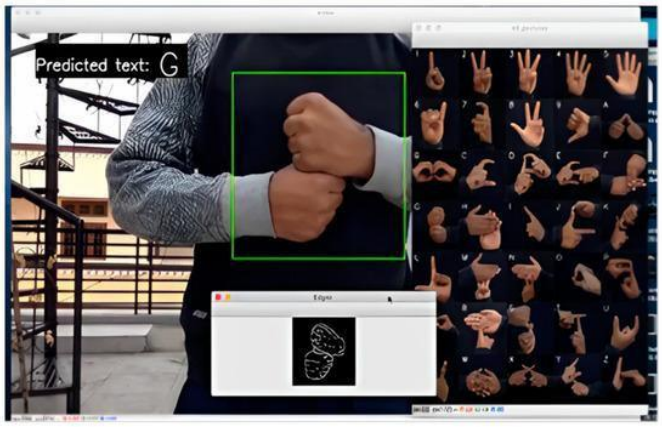

Figure 2 : Output Source

VII. PERFORMANCE METRICS

The performance of the proposed system is evaluated using multiple quantitative metrics to assess its accuracy, efficiency, and real-time capability. These metrics help determine the effectiveness of gesture recognition, sentence generation, and overall system responsiveness.

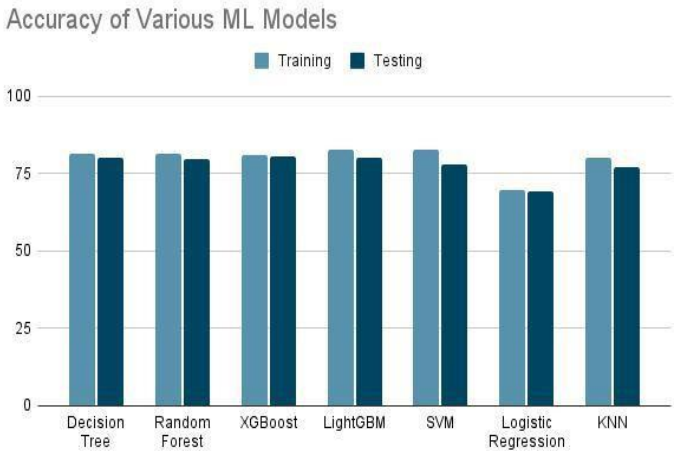

Figure 3:Accuracy Of Various ML Models

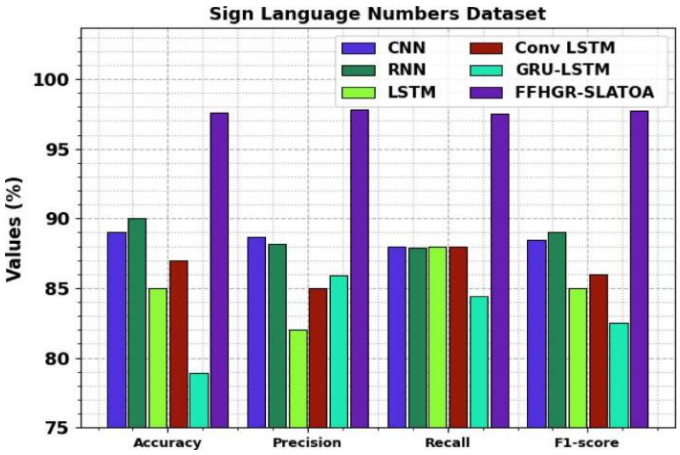

Figure 4: Sign Language Numbers Dataset

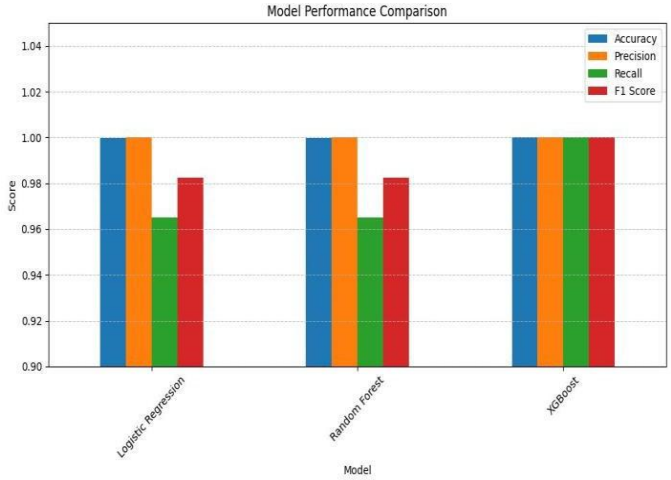

Figure 5 : Model Performace Comparison

Performance Comparison with Existing Methods

Table 2: Performance Comparison with Existing Methods

Method/ System	Gesture Type	Communication Type	Accuracy	Real-Time Support	Limitation
Indian Sign Language Research and Training Centre	Static	One-way	Low	No	No real- time translation
SignA ll	Dynamic	One-way	Medium	Partial	Focused on ASL, not ISL
Google Live Transcribe	No Gesture	One-way	High (Speech)	yes	No sign recognition
Handtalk	Avatar-based	One-way	Medium	Partial	Not ISL, no reverse communication

VIII. RESULT

The proposed AI-based two-way communication system for Indian Sign Language (ISL) demonstrates effective and reliable performance in real-time environments. The integration of MediaPipe enables accurate extraction of hand landmarks, while the Long Short-Term Memory (LSTM) model successfully recognizes dynamic gesture sequences with high accuracy. The system efficiently converts gestures into meaningful text and speech, and also translates speech into text, ensuring seamless bidirectional communication.

Experimental results indicate that the system performs well under normal lighting conditions with minimal latency, making it suitable for practical use. The Natural Language Processing module enhances sentence formation, producing grammatically correct and contextually relevant outputs. Compared to existing methods, the proposed system achieves improved accuracy, faster response time, and better usability.

However, slight performance variations are observed under poor lighting conditions and complex backgrounds. Despite these challenges, the system proves to be a scalable, cost- effective, and user-friendly solution for bridging communication gaps between hearing-impaired and non-hearing individuals.

IX. CONCLUSION

This research presents an AI-based two-way communication system designed to bridge the gap between hearing-impaired individuals and the general population using Indian Sign Language (ISL). By integrating Computer Vision, Deep Learning, and Natural Language Processing techniques, the system enables real-time translation of gestures into text and speech, as well as speech into text. The use of MediaPipe for precise hand landmark detection and Long Short-Term Memory (LSTM) for dynamic gesture recognition ensures high accuracy and efficient performance.

The system demonstrates significant improvements over existing solutions by supporting bidirectional communication, handling continuous gestures, and providing multilingual capabilities. Experimental results confirm that the proposed approach is reliable, scalable, and suitable for real-world applications.

Although certain limitations such as sensitivity to lighting conditions and complex gestures exist, the overall system provides a practical and cost-effective solution. This work contributes toward enhancing accessibility and

promoting inclusivity, paving the way for future advancements in intelligent assistive communication technologies.

X. FUTURE WORK

The proposed system can be further enhanced to improve accuracy, scalability, and real-world applicability. One major direction is the expansion of the dataset to include a wider variety of Indian Sign Language (ISL) gestures, different hand orientations, and diverse environmental conditions, which will help improve model robustness. Future work can also focus on integrating advanced deep learning models such as Transformer-based architectures to enhance contextual understanding and sentence generation.

Additionally, the system can be developed into a mobile application to increase accessibility and usability in everyday scenarios. Improving performance under challenging conditions, such as low lighting and complex backgrounds, is another important area of enhancement. The integration of facial expression recognition can further enrich communication by capturing emotional context.

Moreover, real-time deployment using optimized and lightweight models can make the system more efficient on low- resource devices. Future improvements may also include support for more regional languages and continuous learning mechanisms to adapt to new gestures and user variations, making the system more intelligent and user-centric.

REFERENCES

[1] Camgoz, N. C., Koller, O., Hadfield, S., & Bowden,R. (2020). Sign language transformers: Joint end-to- end sign language recognition and translation. Proceedings of IEEE/CVF CVPR, pp. 10023–10033.

[2] Prakash, N., Kumar, A., & Sharma, R. (2022). Indian sign language recognition using MediaPipe and LSTM. International Journal of Advanced Computer Science and Applications (IJACSA), 13(5), 112–119.

[3] Luqman, H., & Mahmoud, S. A. (2019). Automatic translation of Arabic text-based sign language. Computers & Electrical Engineering, 74, 27–45.

[4] Mohit, B., Naeini, M., Hajishirzi, H., & Smith, N. A. (2022). IndicTrans2: Towards high-quality translation for all 22 scheduled Indian languages. arXiv preprint arXiv:2305.16307.

[5] AI4Bharat. (2023). IndicTTS: Multilingual TTS for Indian languages. https://ai4bharat.iitm.ac.in

[6] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. Advances in NeurIPS, 30.

[7] Lewis, M., Liu, Y., Goyal, N., et al. (2020). BART: Denoising sequence-to-sequence pre-training for NLG, translation, and comprehension. Proceedings of ACL 2020, pp. 7871–7880.

[8] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

[9] Selvaraj, P., & Annamalai, S. (2021). Deep learning-based sign language recognition: A review. Artificial Intelligence Review, 54, 3301–3351.

[10] Adaloglou, N., Chatzis, T., Papadopoulos, G. T., et al. (2022). Comprehensive study on deep learning methods for sign language recognition. IEEE Transactions on Multimedia, 24, 1750–1762.

[11] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT 2019, pp. 4171–4186.

[12] Huang, J., Zhou, W., Li, H., & Li, W. (2018).Attention-based 3D-CNNs for large-vocabulary sign language recognition. IEEE Transactions on Circuits and Systems for Video Technology, 29(9), 2822–2832.

[13] Grindrod, J. (2024). Large language models and linguistic intentionality. Synthese, 204(2), 71. https://doi.org/10.1007/s11229-024-04704-8

[14] Bai, Y., Jones, A., Ndousse, K., et al. (2022). Training a helpful and harmless assistant with reinforcement learning from human feedback. arXiv preprint arXiv:2204.05862.

[15] Koller, O., Zargaran, O., Ney, H., & Bowden, R. (2015). Deep sign: Hybrid CNN-HMM for continuous sign language recognition. Proceedings of BMVC.