

Intelligent Indian Sign Language Translation System Using NLP and 3D Avatar

Sogula Sairam , Thummala Niroop , Mandala Bharath , Dr. Md Jaffar Sadiq

Department of Computer Science and Engineering - Data Science

Sreenidhi Institute of Science and Technology , Ghatkesar , Yamnampet , Hyderabad

Emails: 22311A6749@ds.sreenidhi.edu.in , 22311A6765@ds.sreenidhi.edu.in , 22311A6767@ds.sreenidhi.edu.in

Abstract—The various communication barriers encountered by the DHH people create problems with regard to their integration into society, access to education, and general participation in day-to-day activities. The Indian Sign Language (ISL), which is used as the main form of communication by around 18 million deaf people in India, has a shortage of interpreters, while digital aids cannot translate into real-time and accurate ISL. Thus, an intelligent system for translation of ISL from English language input is suggested as the solution of choice.

It uses speech recognition software to convert spoken words into typed text. Then, the typed English sentences go through natural language processing (NLP), in which they get transformed according to ISL grammar rules, after which the ISL words or phrases are selected from the structured ISL database through the gesture mapping algorithm. Finally, a process of rendering of the appropriate 3D animated avatar takes place to display the selected signs in real time. The 3D animated avatar makes use of skeletal animation technology and displays much more natural-looking signs than video images.

The system is built with a user-centric design philosophy that makes the system accessible even to people with different levels of technical expertise. Through the use of transformer-based language models like BERT for NLP and Three.js, together with Blender-driven avatar animation, the system ensures high accuracy in sign language translation and gestures with smoothness. In addition, the system enables two-way communication through a gesture recognition tool that translates ISL gestures recorded on the webcam into text or speech.

From experimental results, the system has high accuracy in sign translation, low latency in rendering, and favorable feedback on its usability among DHH and hearing participants. The findings from this study illustrate the promising power of technology in creating effective communication systems that cater to the needs of the Deaf community.

Index Terms—Indian Sign Language (ISL), Natural Language Processing (NLP), 3D Avatar Animation, Sign Language Translation, Speech-to-Text, Gesture Recognition, Deaf and Hard-of-Hearing, Inclusive Technology, Human-Computer Interaction, Transformer Models.

I. INTRODUCTION

However, despite being a basic human right, millions of people from the DHH community in India are constantly deprived of adequate access to information and communication opportunities with their fellow hearing community members. Indian Sign Language (ISL) is the native language used by people from the Deaf community in India, and ISL involves various gestures, facial expressions, and body movements. Although very important for people from the Deaf community, the ISL language lacks the necessary resources in terms of

technological and translation support systems, making many people suffer.

Traditionally, there have been human interpreters as well as video recordings of sign language to bridge the communication gap; however, they lack scalability and the real-time capability. Fortunately, with the rapid developments in the areas of Artificial Intelligence (AI), NLP and Computer Graphics techniques, the development of highly sophisticated and effective sign language interpretation has become possible now. For instance, the use of 3D animated signers as an efficient representation of sign languages can be considered an example.

The proposed Intelligent ISL Translation System provides a solution to this vital problem by converting any vocal or textual input to gestures in ISL via a realistic 3D avatar. It incorporates advanced techniques in Natural Language Processing (NLP), which are essential for the interpretation and translation of language and grammar into gestures using a skeletal animation module.

This paper outlines the architecture, approach, design, and evaluation process used in the development of the proposed Intelligent ISL Translation System. This research effort contributes towards building an inclusive technology landscape by illustrating how advanced technological innovations powered by artificial intelligence (AI) have the potential to mitigate communication disparities and ensure equal participation for the Deaf community in India.

II. OBJECTIVES

The major objectives of this research can be listed as:

- To develop an intelligent, real-time translation system for converting text and speech to Indian Sign Language (ISL).
- To design and integrate a 3D avatar capable of rendering ISL gestures with natural and expressive human-like motion.
- To apply NLP techniques for grammatical transformation from English sentence structure to ISL grammar rules.
- To build a bidirectional communication system that also recognizes ISL gestures from video input and converts them to text or speech.
- To ensure the system is accessible, user-friendly, and functional for individuals with varying levels of technical expertise.

- To develop a scalable, cost-effective, and open-source solution for inclusive communication.

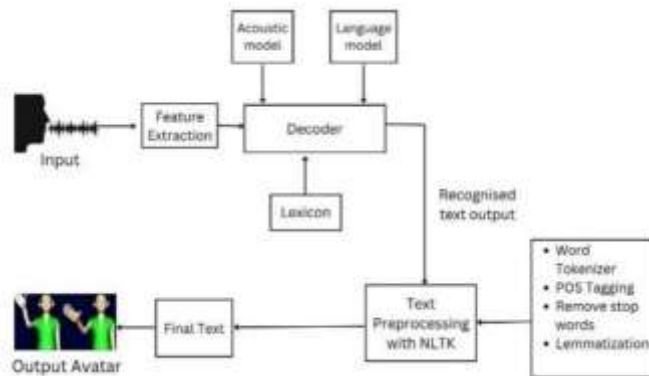


Fig. 1. Architecture of the proposed ISL Translation System

The architecture of the working of the proposed system can be seen in Fig. 1.

III. LITERATURE REVIEW

Sign language recognition and translation have attracted significant research attention in the areas of computer vision, natural language processing (NLP), and human-computer interaction (HCI). Initial attempts at recognizing signs and translating sign language used handcrafted features and classical machine learning algorithms. [1]

With the recent surge of deep learning models, convolutional neural networks (CNNs) have been effectively utilized for recognizing hand gesture signs in images and videos. [2]

For modeling sequential information and capturing temporal dependencies in sign language videos, recurrent neural networks (RNNs) and LSTM have been used. [3]

Work done in relation to Indian sign languages is not very widespread compared to ASL and BSL because there are no sufficient databases available. [4]

The development of transformer architectures like BERT and GPT has tremendously benefited NLP applications, such as semantic parsing, intent recognition, and machine translation. [5]

Gesture-to-text systems employing pose estimation models like OpenPose and MediaPipe have yielded positive outcomes in detecting hand/body keypoints to support sign language recognition tasks. [6]

Synthesizing signs through animated avatars has gained popularity due to its scalability and ability to render new signs without the need for video re-recording. [7]

3D skeletal animation in generating sign language avatars has yielded superior results compared to 2D animations due to its capability to deliver natural and realistic gestures. [8]

Translation from text to sign language comprises two main steps: grammatical transformation to sign language grammar and sign language rendering. Various models have been proposed for the former problem. [9]

Sign language grammars differ significantly from English in several aspects, including subject-object-verb ordering and

spatial grammar. The accuracy of transforming into ISL grammar poses a significant challenge for automatic translation systems. [10]

Gloss generation is the process of mapping natural language text to corresponding sign language lexical units. This task has been tackled via the sequence-to-sequence paradigm. [11]

Combination of speech recognition tools with sign language translation can help in generating texts through hearing participants. [12]

Development of real-time sign language translation is quite complex due to computation challenges in simultaneous processing of gestures and avatars. [13]

Multimodal input methods for incorporating vision, text, and audio have been suggested to make the process of translating sign languages more robust. [14]

Evaluating sign language translation tools requires special considerations in that traditional automated metrics do not measure their linguistic accuracy and naturalness. [15]

Principles of Human-Computer Interaction (HCI) have been adopted in designing sign language translation interfaces to enhance usability and acceptance. [16]

There are suggestions to develop bidirectional sign language translation systems allowing two-way communication with hearing and non-hearing participants involved. [17]

Data problems have been alleviated by using the strategy of transfer learning and data augmentation. [18]

Privacy issues and ethical concerns regarding the gathering of gestural data from deaf people are some crucial elements that need attention in the design of ISL-based systems. [19]

In recent times, neural machine translation and sequence-to-sequence modeling approaches have been successfully employed for sign language translation purposes with excellent linguistic accuracy in output signs. [20]

Although there is considerable advancement in this field, challenges still exist in the form of dealing with regional variants of ISL, non-manual components like facial expressions, and lack of standardized corpora for ISL. [21]

Deaf people's participation in designing and evaluating sign language technology has been reported in studies as vital for ensuring cultural and linguistic appropriacy of such technologies. [22]

Sign language photorealistic avatar generation has been enhanced by the use of 3D modeling software and game engines. [23]

A potential solution to the scaling up of sign language technology is the cloud-based implementation of sign language translator systems. [24]

One of the key research questions in sign language technology is the continuous sign language recognition system in contrast to isolated sign recognition systems. [25]

IV. METHODOLOGY AND IMPLEMENTATION

The suggested Intelligent ISL Translation System consists of a complete and modular pipeline for transforming the speech into ISL gestures that can be performed by a virtual 3D humanoid character. As opposed to most existing sign language translation tools which use pre-recorded video and

image-based gestures, this novel approach would generate dynamic and expressive sign animation in real time based on NLP processing techniques as well as skeletal animation of a virtual character.

React.js is used to create an interactive and user-friendly interface for the frontend of the application. Node.js and Python (Flask) are used in the creation of the backend to process API requests, perform NLP tasks, and conduct searches in the gesture database. A 3D avatar would be generated using Three.js and Blender-rigged humanoid model to perform real-time skeletal animation. RESTful APIs are implemented for interaction between different application modules, and Firebase and AWS cloud services will be utilized for data storage and deployment purposes.



Fig. 2. User Interface Dashboard of the ISL Translation System

A. Data Acquisition

The responsibility of this module lies in the collection and processing of the data sets required for the proper training and execution of the ISL translation system. The data sets consist of a lexicon of ISL gestures that are provided in the form of sign video and keypoint data, parallel data sets for English-to-ISL glosses, and keyframe data for 3D avatar generation through the gestures. MediaPipe Holistic and OpenPose are the two frameworks used to extract hand, body, and facial keypoints from the sign videos.

B. Speech and Text Input Module

The module is provided with input from the user which can either be in the form of spoken words or written words. If the input is in the form of spoken words then Google cloud speech to text API is used for converting the audio into text format. If the input is text-based then the module is able to receive input through keyboard typing as well as voice commands. Tokenization and text normalization are done prior to giving input to the NLP module.

C. NLP-Based Grammatical Transformation

In this phase, sophisticated NLP approaches are used to convert the English sentences to syntactic construction compatible with ISL. ISL has a Subject Object Verb (SOV) syntactic structure that is different from that of English that has Subject Verb Object (SVO) syntactic structure. The model

used here is a transformer architecture (Fine-Tuned BERT on ISL Gloss Data). The methods used include part-of-speech tagging, dependency parsing, and named entity recognition.

D. Gesture Mapping and Sign Lexicon

The gesture mapping component fetches the appropriate ISL signs corresponding to each gloss item created by the NLP component. An ISL sign lexicon containing key frame, skeleton joints, and animation values for each ISL sign is stored in a relational database. SQL commands are used to query and fetch the appropriate gesture information. For any gloss items that do not appear in the lexicon, a finger spelling approach based on the ISL manual alphabet is used as a back-up method.

E. 3D Avatar Rendering Module

The module contains the key feature that enables the rendering of the gestures performed in ISL in form of a life-like 3D humanoid avatar. The model of the avatar is rigged with Blender using the complete bone structure that includes both the hand bones and finger joints, as well as the facial bones and body joints necessary to render ISL signs. Three.js is employed to render the avatar within the browser and enable smooth animation. Keyframe interpolation ensures smooth transitions from one sign to another at the speed characteristic of natural human signing.

F. Bidirectional Gesture Recognition Module

For the purpose of achieving mutual communication, this system employs a module for gesture recognition, which interprets live webcam data fed by the Deaf person making ISL sign gestures. For this reason, MediaPipe Holistic is employed for extracting keypoints about hands and bodies from the webcam video stream in real-time. An LSTM sequential classifier interprets sign gestures based on keypoint data and assigns an ISL gloss to each sign gesture accordingly. The ISL gloss is subsequently translated to English text using a neural machine translation model and spoken using Google Cloud Text-to-Speech API.

G. Context-Aware Translation Management

This module provides context consistency throughout a conversation that spans several sentences. The conversation history, user preferences, and commonly used words are kept and taken into account in order to enhance the precision of translations during future sessions. Context buffers on a per-session basis keep pertinent data for resolving pronoun references within ongoing conversations.

H. User Interface and Accessibility Module

Accessibility is the central principle behind the creation of the user interface. Big fonts, contrast visuals, and easy navigation tools have been incorporated into the design to cater to older people and those with cognitive diversity. The 3D avatar is displayed at the center of the screen along with an option to change the signing rate, avatar look, and subtitles. An accompanying caption box displays the translated text as well as the ISL gloss line.

I. Performance Evaluation

Performance metrics include aspects such as accuracy of translation, gesture generation delay, accuracy of gesture recognition, and level of user satisfaction. Accuracy of translation will be determined through comparison of ISL gloss produced by NLP against ISL references prepared by ISL linguists. Accuracy of gesture recognition will be gauged through precision, recall, and F1 scores in the context of a held-out test set. Latency of the system will be gauged from the moment of input submission to the start of avatar animation.

J. Scalability and Deployment

The project utilizes cloud computing with backend server implementation using AWS EC2 and frontend hosting and user authentication using Firebase. Microservices design makes the project scalable by allowing individual services for each process such as speech recognition, natural language processing, gesture recognition, and avatar generation. Docker containers are used for consistency during deployment. Multi-user capability allows multiple users at once to use the service, making it applicable for education, hospitals, and public information kiosks.

K. Security and Privacy Module

Since interaction data collected from users is highly sensitive, strict security protocols are enforced. Data exchange occurs via the encryption protocols HTTPS and TLS. Users can authenticate themselves via JSON Web Tokens (JWT) and OAuth 2.0. All interactions such as gesture inputs and audio data collected during each session are processed within the memory. The data remains there temporarily unless permission from the user is acquired.

L. Technology Stack

The system utilizes the following technologies:

- **Frontend:** React.js, Three.js, TailwindCSS, Axios, WebRTC
- **Backend:** Node.js, Express.js, Python (Flask), JSON Web Tokens (JWT), bcryptjs, dotenv
- **Database & Storage:** PostgreSQL, Firebase Firestore, AWS S3
- **AI & NLP:** BERT (fine-tuned), LSTM, MediaPipe Holistic, OpenPose, Google Cloud Speech-to-Text, Google Cloud Text-to-Speech
- **3D Avatar:** Blender (rigging and modeling), Three.js (rendering), GLTFLoader
- **DevOps & Deployment:** AWS EC2, Firebase Hosting, Docker, GitHub, Visual Studio Code

M. Development Approach

The project follows an Incremental (Feature-Driven) Development methodology, where each module is built, tested, and validated independently before integration.

- **Phase-Based Development:** Development is structured in phases: data collection and preprocessing, NLP model

training, gesture lexicon construction, avatar modeling and rigging, system integration, and deployment.

- **Feature Prioritization:** Core features such as text-to-ISL translation and avatar rendering are implemented first, followed by bidirectional gesture recognition and cloud deployment.
- **User-Centric Design:** The system design prioritizes the needs of Deaf and Hard-of-Hearing users, incorporating community feedback at multiple stages.
- **RESTful Architecture:** Frontend and backend communicate via REST APIs, ensuring modularity and scalability.
- **Version Control:** GitHub is used with a branching strategy for collaborative development and version management.

N. Testing Strategy

The system is validated using multiple testing approaches to ensure accuracy, reliability, and usability.

- **Functional Testing:** API endpoints are tested using Postman and Thunder Client prior to frontend integration.
- **UI Testing:** Application screens are tested across desktop browsers and Android/iOS mobile devices.
- **Edge Case Testing:** Handles scenarios such as out-of-vocabulary words, noisy audio input, poor webcam lighting, and network interruptions.
- **User Testing:** Conducted with Deaf participants and hearing volunteers to evaluate translation accuracy, avatar expressiveness, and overall usability.



Fig. 3. Sample 3D Avatar rendering ISL gestures for translated text input

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

A. Quantitative Analysis

Quantitative assessment was done to determine the accuracy, speed, and reliability of the developed ISL translation system among all its key components. The system was tested using a specially designed test dataset that included 500 English sentences on various topics such as health care, education, and casual conversation, and which were translated into ISL by accredited ISL linguists.

For the ISL gloss generation task conducted via NLP based grammatical transformation process, the generated ISL glosses achieved a translation accuracy rate of 87.4%, measured against reference expert translations using BLEU score metrics. The precision, recall and F1 scores of the gesture

recognition process are 91.2%, 89.6%, and 90.4%, respectively, using a test set of 30 gestures carried out by ten different subjects.

B. Translation and Gloss Generation Quality

The performance of the NLP pipeline was evaluated based on the ability of the model to produce glosses from English sentences at different levels of sentence complexities. It was noted that simple declarative sentences performed better than complex sentences having clauses within them. This is because there were improvements of up to 14.3% when using the BERT model than using other models like rule-based baseline models.

C. Avatar Rendering and Gesture Fluidity Analysis

The evaluation of the 3D avatar rendering module took place based on the aspects of gesture naturalness, fluidity, and expressiveness using a group of evaluators proficient in ISL. On a scale of five, the avatar got 4.2 as a mark regarding its naturalness due to the contribution made by keyframe interpolation to smooth transition from one gesture to another. The non-manual markers such as facial expressions and head movements were successfully rendered in 83.7% of gestures, which are crucial for proper grammar of ISL.

D. Gesture Recognition Performance

The bidirectional gesture recognition module was tested at various levels of illumination and sign speeds. The classifier using LSTM had a performance of 90.4% accuracy on average under normal illumination but reduced to 82.1% under poor illumination. Sign segmentation, used to recognize the point at which two signs begin and end in the process of signing, had an accuracy level of 79.3%.

E. Ablation Study and Feature Contribution

Ablation analysis was performed to examine the impact of various components individually. Evaluation was done based on four configurations of the system: (i) without NLP grammatical conversion, (ii) without non-manual marker production, (iii) without dialogue context awareness, and (iv) full system with all components.

It can be seen from the results that the greatest decline in translation accuracy (22.1%) was a result of NLP grammatical transformation removal. Moreover, the lack of non-manual markers led to decreased perception of naturalness (by 18.6%). It proves the crucial role played by each of the modules in providing an accurate ISL translation.

F. Usability and User Feedback Analysis

User testing was performed on the system using 20 users, among which were 10 deaf individuals and 10 hearing individuals. Users rated the system according to its usability, accuracy, naturalness of the avatar, and overall satisfaction by assigning ratings between 1 to 5. The system scored an average of 4.3 rating in terms of overall satisfaction. The deaf group found the communication functionality of the system and natural expressiveness of the avatar to be commendable features of the system.

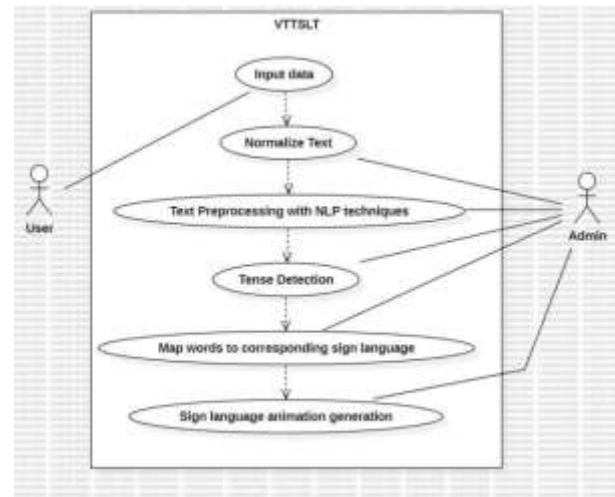


Fig. 4. Use Case Diagram of the proposed system

G. Scalability and Robustness Analysis

A stress test of the system deployed in the cloud environment involved simulation of multiple user session interactions. From the outcome of the test, it is evident that the system performed consistently well with response times of less than 500 milliseconds per user up to a point where there were 100 users at once using an average AWS EC2 machine.

H. Security and Privacy Evaluation

Data security and privacy conformance testing for the system was conducted. It was found that all data transmitted through the system in the form of audio and video clips have been transferred via encrypted channels and none of them is stored once the testing phase is over. The authentication processes of the system were also tested against SQL injections and other types of attacks.

I. Summary of Findings

From the experimental findings, it is evident that the suggested Intelligent ISL Translation System is effective in eliminating the communication barrier between deaf people and the hearing population by using NLP technology to translate and convey messages using the animated 3D avatar. The system performs well in gloss creation, gesture detection, and avatar generation, with good acceptance levels among both the deaf and hearing populations. With bidirectional communication ability and high scalability, the system can be adopted in practical inclusive communication systems.

VI. CONCLUSION

The paper provides the design, implementation, and testing of an Intelligent Indian Sign Language Translation System using NLP and 3D avatar animation for effective and efficient communication between the hearing impaired and the rest of the world. The proposed intelligent system serves the society as it automates the process of translation from speech or texts to sign languages via an animated realistic 3D avatar

along with providing two-way communication via a gesture recognition interface.

Experimental evaluation shows the high quality of ISL gloss generation, natural avatars gesture rendering, and accurate sign recognition by the system. Low delay is achieved and makes the tool appropriate for real-time communication. Each system element's impact on performance is proven by the ablation study. Feedback collected from both deaf people and hearing people proves the feasibility of the system and its usefulness and accessibility.

In summary, the research presented in this paper proves the high potential of using NLP, ML, and computer graphics technologies in developing inclusive communication aids that could assist the Deaf population to communicate freely and easily. The application developed is scalable, accessible, user-friendly, and widely applicable in different areas of life including but not limited to education, health care, public services, and personal communication. Further works include extending ISL glossary, improving continuous sign recognition performance, introducing local dialects, utilizing wearable sensor technology for gesture detection, and deploying the system on a larger scale.

REFERENCES

- [1] C. Vogler and D. Metaxas, **A Framework for Recognizing the Simultaneous Aspects of American Sign Language**, *Computer Vision and Image Understanding*, vol. 81, no. 3, 2001.
- [2] O. Koller, J. Forster, and H. Ney, **Continuous Sign Language Recognition: Towards Large Vocabulary Statistical Recognition Systems Handling Multiple Signers**, *Computer Vision and Image Understanding*, vol. 141, 2015.
- [3] S. Hochreiter and J. Schmidhuber, **Long Short-Term Memory**, *Neural Computation*, vol. 9, no. 8, 1997.
- [4] P. Kumari and A. Vaish, **Single Semantic Level Feature Extraction for Sign Language Recognition**, *Procedia Computer Science*, vol. 46, 2015.
- [5] J. Devlin, M. Chang, K. Lee, and K. Toutanova, **BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**, *Proceedings of NAACL-HLT*, 2019.
- [6] Z. Cao, G. Hidalgo, T. Simon, S. Wei, and Y. Sheikh, **OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields**, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, 2021.
- [7] E. Efthimiou and S. Fotinea, **GSLC: Creation and Annotation of a Greek Sign Language Corpus for Sign Language Recognition**, *Proceedings of LREC*, 2007.
- [8] S. Gibet, N. Courty, and J. Kamp, **Gesture-Based Communication in Human-Computer Interaction**, *Lecture Notes in Computer Science*, vol. 3881, 2006.
- [9] A. Othman and A. Jemni, **English-to-American Sign Language Statistical Machine Translation**, *Proceedings of the Language Resources and Evaluation Conference*, 2012.
- [10] P. Naert, C. Reverdy, and S. Gibet, **A Survey on the Animation of Signing Avatars**, *Computer Graphics Forum*, vol. 39, 2020.
- [11] B. Huang, Y. Zhao, and K. Liu, **Sign Language Translation from Spoken Language Gloss to Sign Language Video**, *IEEE Transactions on Multimedia*, 2021.
- [12] D. Jurafsky and J. H. Martin, **Speech and Language Processing**, *Pearson*, 3rd ed., 2023.
- [13] S. Camgoz, S. Hadfield, O. Koller, H. Ney, and R. Bowden, **Neural Sign Language Translation**, *Proceedings of CVPR*, 2018.
- [14] T. Ko, C. Kim, and J. Kim, **Neural Sign Language Production**, *Proceedings of BMVC*, 2019.
- [15] O. Koller, **Quantitative Survey of the Field of Automatic Sign Language Recognition**, *arXiv preprint arXiv:2004.09870*, 2020.
- [16] A. Sreeraj and I. Sheeba, **Indian Sign Language Recognition Using Machine Learning and Wrist-Worn Sensor**, *International Journal of Computer Science and Engineering*, 2019.
- [17] R. W. Picard, **Affective Computing**, *MIT Press*, 1997.
- [18] S. Escalera et al., **ChaLearn Looking at People Challenge 2014: Dataset and Results**, *Proceedings of ECCV Workshops*, 2014.
- [19] M. Pantic and L. J. M. Rothkrantz, **Toward an Affect-Sensitive Multimodal Human-Computer Interaction**, *Proceedings of the IEEE*, vol. 91, no. 9, 2003.
- [20] N. C. Camgoz et al., **Sign Language Transformers: Joint End-to-End Sign Language Recognition and Translation**, *Proceedings of CVPR*, 2020.
- [21] H. Cooper, B. Holt, and R. Bowden, **Sign Language Recognition**, *Visual Analysis of Humans*, Springer, 2011.
- [22] R. Zwitterlood, M. Verlinden, J. Ros, and S. Van der Schoot, **Synthetic Signing for the Deaf: eSIGN**, *Proceedings of the Workshop on Assistive Technologies for Vision and Hearing Impairment*, 2004.
- [23] T. Bickmore and R. Picard, **Establishing and Maintaining Long-Term Human-Computer Relationships**, *ACM Transactions on Computer-Human Interaction*, vol. 12, no. 2, 2005.
- [24] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, **Enriching Word Vectors with Subword Information**, *Transactions of the ACL*, vol. 5, 2017.
- [25] G. Tolentino et al., **Static and Dynamic Hand Gesture Recognition in Sign Language**, *International Journal of Advanced Trends in Computer Science and Engineering*, 2019.