Proceedings of the 3rd International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/138/2025.21526

Intelligent Bidirectional Sign Language Translator – Sign Language Communication

Xiangyang Zhang[1,a,*]

[1]Xidian University, Xi'an City, Shaanxi Province, 710126, China
a. 22009200279@stu.xidian.edu.cn
**corresponding author*

Abstract: With the increasing societal focus on the communication needs of individuals with hearing impairments, this paper proposes an intelligent bidirectional sign language translator – Sign Language Communication – aimed at building an efficient bridge for communication between hearing-impaired individuals and hearing individuals through technological innovation. The Sign Language Communication system integrates gesture capture modules, motion perception modules, information projection modules, speech input modules, and speech output modules, enabling functions such as data collection, real-time sign language recognition, bidirectional speech conversion, and visual information display. Research demonstrates that this system has broad applicability in sign language translation scenarios such as daily communication, business meetings, and educational classrooms, as well as in speech recognition-based interactions. Additionally, it caters to personalized user needs, providing hearing-impaired individuals with an efficient and convenient communication solution.

Keywords: Deep learning, Sign language translation, Sign language recognition, Bidirectional communication

1. Introduction

With the rapid development of society and continuous technological advancement, the level of informatization has been significantly enhanced. However, individuals with hearing impairments still face numerous communication barriers in their daily lives and work. Although sign language serves as an effective communication tool, the fact that most hearing individuals are unfamiliar with it often restricts the information exchange and emotional interaction for those with hearing impairments. This challenge is particularly pronounced in scenarios requiring efficient information transmission, such as education and conferences, where hearing-impaired individuals encounter inequities in information access. Such circumstances not only hinder their social integration but also impede the development of an inclusive society. In response, the “Sign Language Translator” system has been designed to leverage intelligent technology to eliminate communication barriers, embody humanistic care for individuals with hearing impairments, and contribute to building a more equitable and harmonious society.

Sign language, a language naturally developed for daily communication between the hearing-impaired and hearing populations, primarily relies on hand gestures and is sometimes supplemented with facial expressions and eye contact. Sign language recognition technology uses computer systems to recognize sign language signals collected via specific sensors or cameras, ultimately converting these signals into text. This process integrates disciplines such as computer vision, pattern recognition, and natural language processing, making it a challenging yet promising research area. The technology exhibits broad application prospects across fields such as education, healthcare, social interactions, human-computer interaction, and intelligent robotics control [1][2].

Research on sign language recognition has accumulated over time, evolving through two primary stages: early traditional methods and the more recent deep learning-based approaches. Each stage has driven advancements in sign language recognition and fueled aspirations to foster barrier-free communication. These developments not only represent the progress within the field but also underscore the pivotal role of technological improvements in enhancing recognition accuracy and broadening application scenarios.

Early sign language recognition primarily relied on handcrafted feature extraction, analyzing gesture geometry, motion trajectories, and spatial positions using predefined rules. Imagawa et al. used local feature clustering to represent hand images as cluster symbols for recognition [3]. Starner and Pentland applied Hidden Markov Models (HMM) to capture temporal gesture features, improving dynamic gesture recognition [4]. Fels and Hinton employed data gloves to collect hand movement data and map them to symbols [5]. However, early methods heavily depended on researchers' deep understanding of sign language features and prior knowledge, resulting in limited applicability, low accuracy, and poor generalizability. Faced with the complexity and diversity of sign language expressions, traditional methods proved insufficient.

With the continuous progress in artificial intelligence and computer vision technologies [6], deep learning has been extensively applied in sign language recognition to overcome the limitations of traditional methods. By autonomously learning data characteristics, deep learning reduces reliance on handcrafted features, markedly improving recognition accuracy, expanding applicability, and enhancing system robustness. Deep learning-based sign language recognition techniques can be categorized into three main types: 2D image-based recognition, video sequence-based recognition, and wrist sensor-based recognition.

2D image-based sign language recognition primarily analyzes gesture features within single-frame images, often using convolutional neural networks (CNNs) to extract key shape and texture information from hand regions. Koller et al. proposed a CNN-based model that improved recognition accuracy by capturing fine-grained features [7]. To mitigate issues like background complexity and lighting variations, researchers [8-11] integrated multimodal data with multi-scale feature extraction techniques, enhancing static gesture recognition. While CNNs effectively learn hand image features, 2D image-based methods struggle with recognizing dynamic sign language sequences.

Video-based sign language recognition compensates for the limitations of single-frame image methods by analyzing dynamic gesture sequences, effectively capturing the continuous motion information of sign language. This approach typically employs recurrent neural networks (RNNs) or long short-term memory networks (LSTMs) to model temporal sequence characteristics and accurately represent gesture variations over time. Pu et al. proposed a method combining 3D CNNs and bidirectional LSTMs to extract spatial and dynamic features within the spatiotemporal domain, significantly improving recognition accuracy and robustness [12]. Yang et al. [13][14] designed a model integrating CNNs and LSTMs, which processes both RGB images and optical flow data through a dual-input structure, outputting classification results via fully connected layers. This lightweight model performed well on small-scale sign language datasets and met real-time processing requirements.

Wrist sensor-based sign language recognition combines deep learning with motion data captured by wearable devices, enabling precise recognition through data collected from wrist movements. Deep learning models autonomously extract complex dynamic features from sensor data, achieving

DOI: 10.54254/2755-2721/138/2025.21526

accurate gesture recognition. Compared to traditional image or video methods, wrist sensor data better captures subtle hand movements (e.g., wrist rotation, acceleration changes) and is unaffected by lighting variations. Matej Králik and Marek Šuppa proposed a system called “WaveGlove,” which combines inertial sensor data from gloves with Transformer models, achieving high-accuracy recognition of complex gestures [15].

Despite significant advancements in sign language recognition research, several limitations persist. Most studies overly focus on single technologies, such as sign language recognition, speech recognition, or speech synthesis, and remain at the theoretical validation stage without practical application. Consequently, while numerous theoretical achievements have been made, these technologies have yet to be effectively transformed into practical, usable products.

To address this issue, the “Sign Language Translator” system integrates existing technologies into a comprehensive solution. Its functionalities encompass the entire process from data collection to sign language translation, speech synthesis, and recognition. The system incorporates core modules such as gesture capture, motion sensing, information projection, speech input, and output, with different modules working collaboratively to achieve real-time sign language recognition, bidirectional speech conversion, and information visualization.

The system not only employs cutting-edge technologies such as smart cameras, sensors, and deep learning algorithms but also emphasizes product implementation and user experience optimization. Unlike previous studies that focused solely on specific domains, the “Sign Language Translator” prioritizes multi-scenario applicability and real-time performance. Research demonstrates that the system provides hearing-impaired individuals with efficient and convenient communication tools across scenarios such as daily interactions, business meetings, and educational settings. Additionally, it offers customized services tailored to user needs, realizing the goal of empowering inclusive societal development through technology.

2. System Module Design

As shown in the figure below, the Sign Language Communication System provides robust support for seamless communication between sign language and spoken language through high-precision modular design. The core modules of the system include the Gesture Capture Module, Motion Perception Module, Information Projection Module, Speech Input Module, and Speech Output Module. These modules work closely together to ensure the system’s accuracy, real-time performance, and adaptability to various scenarios.

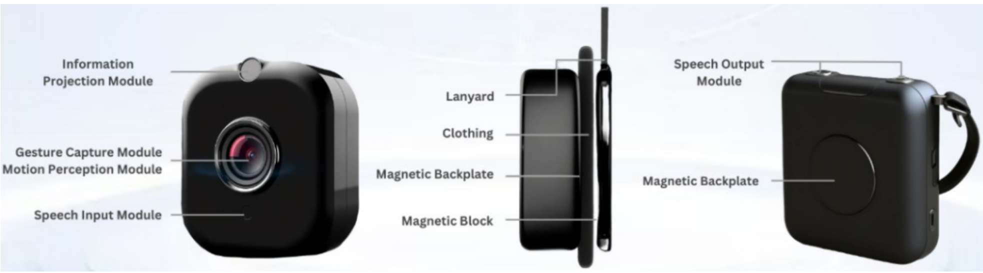

Figure 1: Functional Modules and Product Design Diagram.

2.1. Gesture Capture Module

The intelligent camera plays a crucial role in the Sign Language Communication System. With its high-precision image recognition capabilities, real-time feedback mechanism, and strong adaptability to diverse environmental conditions, it successfully captures and converts sign language gestures into understandable information.

DOI: 10.54254/2755-2721/138/2025.21526

In terms of image recognition and gesture tracking, the high-resolution sensors built into the camera can capture every subtle movement of the user's hands. From minor finger movements to the rotation angles of the palm, these rich gesture details are swiftly captured and deeply analyzed using computer vision algorithms. This process includes precise measurements of gesture position and angle as well as real-time tracking of dynamic information such as gesture duration and directional changes, enabling comprehensive and accurate recognition of complex sign language gestures.

The intelligent camera combines image processing algorithms with deep learning techniques. Using pre-trained neural networks, it processes input images to accurately recognize and interpret gestures, enabling the system to identify various sign language symbols and adapt to both skilled and novice users, broadening its application range.

Additionally, the camera adapts well to different lighting conditions. Whether outdoors, indoors with fluctuating light, or in low-light environments, it uses image enhancement techniques to adjust exposure, contrast, and brightness, ensuring clear images and accurate recognition. This makes the system suitable for various settings, from formal meetings to daily conversations, providing seamless communication.

2.2. Motion Perception Module

Sensors work in coordination with the camera to enhance the system's perceptive capabilities and improve the accuracy of sign language recognition. Utilizing advanced sensing technologies such as infrared, ultrasound, or others, the sensors track spatial displacements and fine hand movements, providing valuable supplemental data to the system.

By capturing the three-dimensional spatial positions of the hands, the sensors offer critical support in interpreting the specific meaning of sign language gestures, particularly for symbols requiring precise finger position adjustments or gesture direction changes. The sensors ensure the system comprehends each gesture detail accurately, minimizing misunderstandings or ambiguities.

The application of inertial sensors such as accelerometers and gyroscopes further elevates the detection of hand dynamics. During rapid movements or intricate gestures, these sensors capture details that the camera might miss and correct any discrepancies in image recognition. Hand rotations, flips, and shakes are accurately identified, effectively reducing errors and enhancing translation accuracy.

Real-time and high-frequency data collection is another key advantage of the sensors. Upon detecting hand movements, the sensors promptly transmit data to the system's processing unit, where it integrates with image information from the camera. This real-time feedback mechanism ensures translation fluency, providing timely and accurate responses to every gesture made by the user.

2.3. Information Projection Module

The core function of the Information Projection Module is to convert received speech inputs into text and project the text onto the user's palm or a designated surface. This enables individuals with hearing impairments to quickly comprehend spoken language visually, thereby breaking communication barriers.

Given that the system operates in varied lighting environments, the projection module's ability to adapt to changing light conditions is crucial. The high-brightness design ensures text remains legible under all lighting conditions. Furthermore, the projection lens features automatic brightness adjustment and light-sensing technology, which modify projection brightness based on ambient light levels to maintain optimal visual clarity across different scenarios. This intelligent design not only enhances the user experience but also ensures effective information delivery.

DOI: 10.54254/2755-2721/138/2025.21526

The module also exhibits multi-scenario adaptability. With adjustable focus and flexible projection angles, users can modify the size and position of the projection through gestures or other means to suit diverse needs. Its high resolution ensures that even small fonts or intricate symbols are displayed clearly, avoiding blurriness or distortion. Additionally, the low-reflection projection design minimizes interference from environmental light, making the projected content more stable and legible.

2.4. Speech Input Module

The Speech Input Module converts spoken input from non-hearing-impaired individuals into text, which is then processed by the system and projected in real time for hearing-impaired users.

This module's core functionality begins with high-quality speech acquisition. Using high-sensitivity microphones and noise suppression algorithms, the system ensures clear capture of the speaker's speech in complex environments, ranging from noisy public spaces to quiet private settings. The microphones employ directional and noise reduction technologies to filter out background noise effectively, while echo cancellation further reduces interference from reflective sounds, guaranteeing clear and stable audio input.

The captured speech signals are transmitted to the Automatic Speech Recognition (ASR) system. Leveraging deep learning models and extensive speech databases, the ASR system converts audio waveforms into recognizable phonemes, words, and sentences, ultimately generating textual output. The process emphasizes both high accuracy and real-time performance, enabling the system to respond promptly to spoken input. To achieve this, the ASR system continually optimizes its algorithms, minimizing latency and improving recognition efficiency. It also utilizes contextual information for intelligent optimization, further enhancing accuracy.

2.5. Speech Output Module

The speech output module converts sign language translation results into speech, allowing non-hearing-impaired individuals to clearly hear the translated content and enabling two-way communication. This process involves converting sign language to text and then using text-to-speech (TTS) technology, which demands high clarity and real-time performance.

The module's core function is TTS technology, which converts text from the sign language translation system into clear, natural speech. The text is derived from sign language movements captured by cameras and sensors, processed by the system, and then synthesized into speech to ensure smooth audio output.

Unlike traditional mechanical speech, modern speech synthesis technologies, such as neural network-based synthesis, make the generated speech sound more natural and expressive. By adjusting tone, speed, and emotional nuances, the system can produce speech output that aligns better with the tone of natural conversation, facilitating smoother interactions between hearing-impaired and non-hearing-impaired individuals. Moreover, the speech output module delivers high-quality audio output, ensuring clarity even in noisy environments. The system can automatically adjust volume and audio quality to maintain intelligibility.

Real-time performance is a crucial feature of the speech output module, requiring low latency to ensure conversational fluidity. The speed of speech synthesis must be fast enough to remain synchronized with the real-time performance of the sign language translation system. Additionally, the module is designed to adapt to different environments, capable of automatically increasing volume or adjusting output modes in noisy conditions to ensure the speech is clearly audible. The audio output must also have excellent sound field coverage, enabling non-hearing-impaired individuals in various positions to hear the speech content clearly.

DOI: 10.54254/2755-2721/138/2025.21526

2.6. Magnetic Design: Portability and Usability

The ingenious combination of a magnetic backplate and a magnetic block allows the device to attach easily and securely to the user's clothing without the need for cumbersome straps or additional carrying tools, significantly simplifying the process of carrying and using the device. This design provides privacy by discreetly attaching to clothing, allowing hearing-impaired users to engage in daily communication confidently without drawing unwanted attention.

Moreover, the magnetic attachment ensures the device remains stable during use, allowing the gesture recognition module to accurately capture hand movements and minimize errors from shaking or misalignment, thus improving translation accuracy. The projection module is also highly flexible, automatically adjusting its angle to project translated text clearly onto the user's hand or other visible areas, ensuring clear text display whether the user is standing or sitting, without the common issues of improper angles or blurry projections.

3. Introduction to Application Scenarios

The "Sign Language Communication System" (SLC System) not only exhibits theoretical completeness in design but also demonstrates extensive adaptability and contextual functionality in practical applications. Through the collaborative operation of multiple modules, the system flexibly accommodates diverse communication needs, meeting the personalized requirements of individuals with special needs. Below is a detailed introduction to several typical application scenarios and their working principles.

3.1. Sign Language Translation Scenario

In modern society, individuals with hearing impairments, despite having a unique sign language communication system, often face barriers in interacting with people who do not understand sign language. In everyday situations, when hearing-impaired individuals attempt to express emotions and needs through sign language, their messages are frequently obstructed by an invisible barrier, making effective communication difficult.

To address this issue, the SLC System was developed. As illustrated in Figure 2, the entire workflow includes key stages such as data collection, feature extraction, algorithm recognition, and speech output. The system's camera, sensors, and speech output module collaborate to create an efficient and accurate sign language recognition and translation system, facilitating the integration of hearing-impaired individuals into modern society.

Hearing-impaired users can easily attach the SLC System to their clothing, particularly in the chest or shoulder area, using magnetic adhesion. The camera, once positioned, remains focused on the user's hand region, capturing even the most subtle movements.

During the data collection phase, an intelligent camera serves as the core component, utilizing built-in high-precision image processing algorithms to capture gesture movements in real time. The camera accurately identifies dynamic changes such as finger bending, palm rotation, and arm extension, extracting key features of complex sign language actions and converting them into digital signals, which provide a reliable data foundation for subsequent processing.

The sensor further improves recognition accuracy by capturing small changes in force, speed, and trajectory in sign language movements. This dynamic information is combined with the data captured by the camera to build a more comprehensive cognitive model of sign language. The fused data will be transmitted to the algorithmic recognition unit, which parses the sign language movements through efficient pattern recognition algorithms, thus significantly reducing the risk of misjudgment and comprehensively improving the recognition accuracy and real-time performance of the system.

DOI: 10.54254/2755-2721/138/2025.21526

Once gesture features are extracted, the system immediately initiates translation, utilizing advanced speech synthesis technology to convert sign language meanings into clear and natural speech. This capability is supported by efficient natural language processing algorithms and robust computational power, ensuring accurate real-time translation. The speech synthesis module also incorporates variations in emotion and tone to deliver context-appropriate vocal expressions. A high-quality speech module ensures clear and smooth audio output, enabling individuals who are not hearing impaired to easily understand the translated content. This facilitates efficient and seamless two-way communication, breaking down barriers between hearing and hearing-impaired individuals.

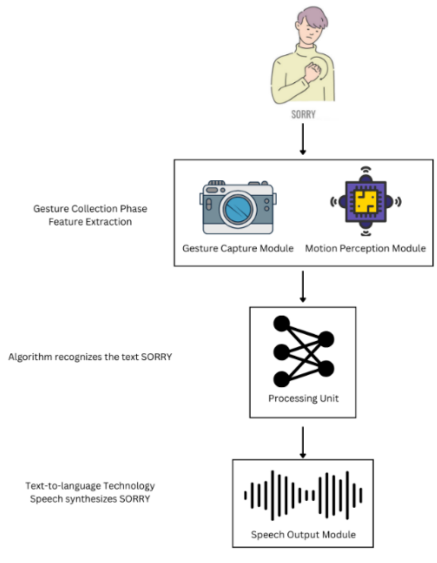

Figure 2: Flowchart of Sign Language Translation.

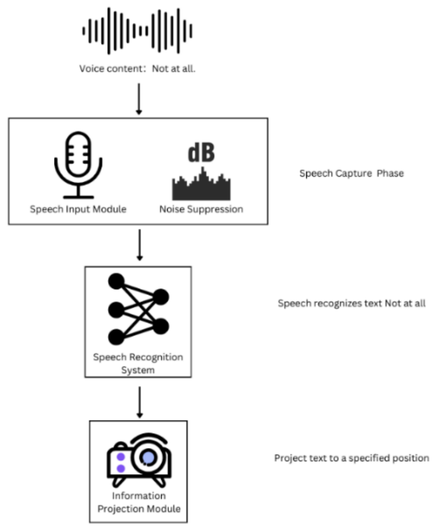

Figure 3: Flowchart of Speech Recognition.

3.2. Speech Recognition and Interaction Scenario

In environments like meetings or classrooms, hearing-impaired individuals often struggle to fully comprehend speech from non-hearing-impaired individuals, impacting their learning and information acquisition. To address this, the SLC System uses a speech input module and an information projection module for efficient speech-to-text conversion, ensuring timely and accurate information delivery. As shown in Figure 3, the workflow includes key stages such as speech signal acquisition, speech recognition, and information projection.

The system uses a high-sensitivity microphone to capture the speaker's speech, integrating noise suppression algorithms for accurate signal capture in various environments. In noisy settings, the system effectively reduces background noise, ensuring clear transmission of speech. The captured signals are sent to the speech recognition system, which quickly converts the spoken content into text.

The translated text is then projected in real time onto the palm of the hearing-impaired individual or a designated area in front of them via the information projection module. This process is seamless, allowing users to keep up with the speaker's content. The display position and font size can be adjusted based on user preferences and specific scenarios for optimal readability.

In addition to real-time speech-to-text conversion, the system includes an information storage function. Speech content and the corresponding translated text from meetings or classes can be saved for future reference. By integrating cloud storage, users can access past communication records as needed, allowing them to better retain and review important information, especially for revisiting critical discussions in learning or work contexts.

DOI: 10.54254/2755-2721/138/2025.21526

3.3. Expansion of Application Scenarios

The SLC system is not only equipped with efficient bidirectional communication functions but also significantly enhances user experience through personalized settings, ensuring that each user can customize their communication methods based on individual needs. The following illustrates how personalized settings are applied across different scenarios:

Personalized Settings for the Projection Module - The projection module of the SLC system allows users to adjust the display of translation content according to their visual needs. These adjustments include font size, color, contrast, and more, ensuring readability and comfort. Through these personalized settings, users can modify the display based on their reading preferences or vision conditions. For instance, some hearing-impaired users may prefer larger fonts, while others might require specific color contrasts to enhance visual clarity. This flexible customization meets diverse needs, ensuring users can easily access information.

Support for Multilingual Functionality - The gesture capture module can translate sign languages from various countries, meeting the global demand for sign language communication. Similarly, the speaker module offers high personalization and intelligence. It supports multiple languages and can adapt to the current linguistic environment, enabling smooth speech switching. For example, in cross-cultural communication, the system can automatically adjust translation based on the language and cultural habits of different users, ensuring smooth communication for hearing-impaired individuals across regions. Multilingual support enhances the system’s adaptability and ensures global applicability.

Personalized Settings for Speech Interaction - In the area of speech interaction, the system incorporates speech command functionality specifically designed for non-hearing-impaired users. This feature allows users to interact seamlessly with the system using natural language while enabling adjustments to speech speed and tone according to personal preferences or auditory conditions. This ensures that every speech output aligns precisely with the user’s auditory habits and communication expectations.

Customization of Sign Language Gestures - In sign language translation scenarios, the system allows users to add or modify personal sign language gestures via a companion app, ensuring accurate recognition of individual expressions. Whether due to regional variations or specific user needs, the system adapts through customizable settings. This feature enables the SLC system to transcend geographical and cultural boundaries, acting as a seamless communication bridge for hearing-impaired communities globally.

These personalized settings enhance the system's adaptability to diverse linguistic and cultural contexts, while also improving its humanized service, significantly boosting user comfort and interactivity.

4. Experiment

To validate the effectiveness of the SLC system in real-world communication, this paper designed an experimental scenario simulating an everyday conversation between a hearing-impaired individual and a hearing individual. The experiment involved one hearing-impaired participant and one hearing participant, who engaged in a conversation both with and without the assistance of the system, then observed differences in communication fluency, comprehension accuracy, and user experience.

DOI: 10.54254/2755-2721/138/2025.21526

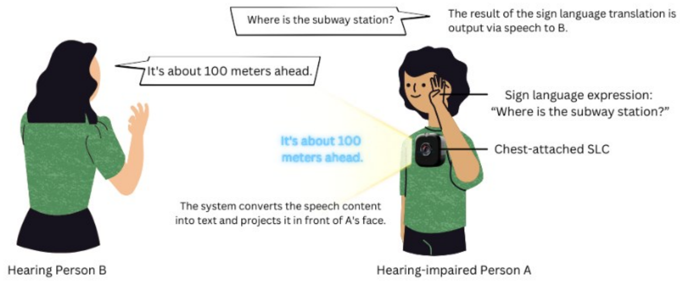

Figure 4: Schematic Diagram of a Communication Scenario.

4.1. Scenario Description

In the experiment, the hearing-impaired individual (A) and the hearing individual (B) met on the road, where A wanted to ask B for directions to the nearest subway station.

Without the System: A attempted to communicate using hand gestures and by typing on a mobile device. However, B, who was unfamiliar with sign language, struggled to understand the gestures and had to wait for A to type out the message. This led to slow communication and potential misunderstandings.

With the System: As shown in Figure 4, A used the SLC system to facilitate the conversation. The intelligent camera and sensors captured A's hand gestures in real-time, and the deep learning algorithm translated them into text, which was then converted into speech for B to hear. In response, B spoke their answer, and the system transcribed the speech into text, projecting it onto A's palm or onto the table in front of them. This enabled A to read and understand B's response instantly.

4.2. Experimental Results and Analysis

The results showed that with the assistance of the SLC system, communication between the two individuals became significantly more efficient. A could express their question quickly, while B could understand and respond seamlessly. Compared to unaided communication, the system reduced transmission delays, improved comprehension accuracy, and enhanced the overall interaction experience.

5. Conclusion

This paper presents an innovative intelligent bidirectional sign language translation system—"Sign Language Communication"—designed to provide an efficient and convenient solution for communication between hearing-impaired individuals and those with normal hearing. By integrating multiple advanced technologies, the system establishes a comprehensive architecture encompassing data collection, sign language translation, and speech synthesis and recognition.

The core modules and functional implementations of the "Sign Language Communication" system are introduced in detail, including the gesture capture module, dynamic perception module, information projection module, speech input module, and speech output module. The gesture capture module uses intelligent cameras and sensors to accurately record and track sign language gestures in real time. The dynamic perception module extracts motion features and combines them with deep learning algorithms to improve the accuracy and robustness of sign language recognition. The information projection module displays the converted text or images clearly within the user's visible range, ensuring the intuitive transmission of information. The speech input module captures speech information through highly sensitive microphones and converts it into text, while the speech output module uses speech synthesis technology to provide natural speech output of sign language content, fulfilling the bidirectional communication needs of hearing-impaired and hearing individuals.

The paper also analyzes the typical application scenarios of the “Sign Language Communication” system. Research shows that the system can provide efficient translation support in various settings, such as daily communication, business meetings, and educational classrooms, helping hearing-impaired individuals integrate seamlessly into social activities. It also pays special attention to users’ personalized needs, significantly enhancing user satisfaction through meticulous functional design. By integrating multiple advanced technologies and focusing on product design, the “Sign Language Communication” system demonstrates remarkable advantages in real-time performance, accuracy, and applicability, further promoting the realization of inclusive social development.

References

[1] *Gao, W., Chen, X. L., Ma, J. Y., & Wang, Z. Q. (2000). A communication system for deaf and normal people based on multi-modal interface technology. Journal of Computer Research and Development, (12), 1253–1260.*

[2] *Ong, S., & Ranganath, S. (2005). Automatic Sign Language Analysis: A Survey and the Future beyond Lexical Meaning. IEEE transactions on pattern analysis and machine intelligence. 27. 873-91. 10.1109/TPAMI.2005.112.*

[3] *Imagawa, H., Matsuo, R., Taniguchi, D., Arita, S., Lu, S., & Igi, S. (2000). Recognition of local features for camera-based sign language recognition system. In Proceedings of the 15th International Conference on Pattern Recognition (ICPR-2000) (Vol. 4, pp. 849–853). Barcelona, Spain. https://doi.org/10.1109/ICPR.2000.903050*

[4] *Starner, T., Weaver, J., & Pentland, A. (1998). Real-time American sign language recognition using desk and wearable computer-based video. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(12), 1371–1375. https://doi.org/10.1109/34.735811*

[5] *Fels, S. S., & Hinton, G. E. (1993). Glove-Talk: A neural network interface between a data-glove and a speech synthesizer. IEEE Transactions on Neural Networks, 4(1), 2–8. https://doi.org/10.1109/72.182690*

[6] *Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. Neural Computation, 18(7), 1527–1554. https://doi.org/10.1162/neco.2006.18.7.1527. PMID: 16764513.*

[7] *Koller, O., Zargaran, O., Ney, H., & Bowden, R. (2016). Deep Sign: Hybrid CNN-HMM for Continuous Sign Language Recognition. The British Machine Vision Conference(BMVC)2016, York.https://openresearch.surrey.ac.uk/esploro/outputs/conferencePresentation/Deep-Sign-Hybrid-CNN-HMM-for-Continuous/99511412402346#file-0*

[8] *Tang, Ao & Lu, Ke & Wang, Yufei & Huang, Jie & Li, Houqiang. (2015). A Real-Time Hand Posture Recognition System Using Deep Neural Networks. ACM Transactions on Intelligent Systems and Technology. 6. 1-23. 10.1145/2735952.*

[9] *Pigou, L., Dieleman, S., Kindermans, PJ., Schrauwen, B. (2015). Sign Language Recognition Using Convolutional Neural Networks. In: Agapito, L., Bronstein, M., Rother, C. (eds) Computer Vision - ECCV 2014 Workshops. ECCV 2014. Lecture Notes in Computer Science(), vol 8925. Springer, Cham. https://doi.org/10.1007/978-3-319-16178-5_40*

[10] *Kang, B., Tripathi, S., & Nguyen, T. (2015). Real-time sign language fingerspelling recognition using convolutional neural networks from depth map. 136-140. 10.1109/ACPR.2015.7486481.*

[11] *Hossen, M. A., Govindaiah, A., Sultana, S., & Bhuiyan, A. (2018). Bengali sign language recognition using deep convolutional neural networks. In Proceedings of the 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR) (pp. 369–373). Kitakyushu, Japan. https://doi.org/10.1109/ICIEV.2018.8640962*

[12] *Pu, J., Zhou, W., & Li, H. (2016). Sign language recognition with multi-modal features. In E. Chen, Y. Gong, & Y. Tie (Eds.), Advances in Multimedia Information Processing – PCM 2016 (Vol. 9917, pp. 281–292). Cham: Springer. https://doi.org/10.1007/978-3-319-48896-7_25*

[13] *Su, Y., & Zhu, Q. (2017). Continuous Chinese sign language recognition with CNN-LSTM. In Proceedings of the Ninth International Conference on Digital Image Processing (ICDIP 2017). https://doi.org/10.1117/12.2281671*

[14] *Yang, S., & Zhu, Q. (2017). Video-based Chinese sign language recognition using convolutional neural network. In Proceedings of the 9th IEEE International Conference on Communication Software and Networks (ICCSN) (pp. 929–934). Guangzhou, China. https://doi.org/10.1109/ICCSN.2017.8230247*

[15] *Králik, M., & Šuppa, M. (2021). Waveglove: Transformer-based hand gesture recognition using multiple inertial sensors. In Proceedings of the 29th European Signal Processing Conference (EUSIPCO) (pp. 1576–1580). Dublin, Ireland. https://doi.org/10.23919/EUSIPCO54536.2021.9616000*