تحويل الصوت إلى إشارة يدوية: نظام تواصل مساعد لضعاف السمع

د/ عبير محمد حسن سعد
مدرس بقسم إعداد معلم الحاسب الآلي
كلية التربية النوعية – جامعة دمياط

د/ دعاء محمد أبوراشد أمين هوى
مدرس بقسم إعداد معلم الحاسب الآلي
كلية التربية النوعية – جامعة دمياط

المستخلص:

يهدف هذا البحث إلى مساعدة الأفراد الذين يعانون من ضعف السمع من خلال تطوير نظام يحول الكلمات المنطوقة إلى لغة إشارة. تضمنت المنهجية تحويل المدخلات الصوتية إلى إشارة باستخدام خوارزمية التعرف على الكلام، والتي تم تحويلها بعد ذلك إلى إيماءات لغة الإشارة الأمريكية (ASL) المقدمة في الوقت الفعلي. تم استخدام أدوات مثل التعرف على الكلام وواجهة برمجة تطبيقات Google للنسخ الدقيق والإخراج المرئي. وتؤكد النتائج والتحليلات على جدوى نظام تحويل الصوت إلى لغة الإشارة كأداة مساعدة للأشخاص الذين يعانون من ضعف السمع.

الكلمات المفتاحية : تحويل الصوت إلى لغة الإشارة– تمثيل لغة الإشارة– لغة الإشارة الأمريكية– التواصل في الوقت الحقيقي– واجهة المستخدم الرسومية– معالجة اللغة الطبيعية.

Voice-to-Hand Sign Transformation: An Assistive Communication System for the Hearing Impaired

Abstract:

This research aims to assist individuals with hearing impairments by developing a system that converts spoken words into sign language. the study was conducted with a sample of five deaf participants who evaluated the system . the methodology involved transforming audio input into signal using speech recognition algorithm , which were then converted into American sign language (ASL) gestures presented in real-time. Tools such as speech recognition and Google's API were used for accurate transcription and visual output. The system effectiveness was assessed through participant feedback and accuracy measurements, showing promising results. The research concludes with a

recommendation for further enhancement of the system accuracy and user experience to better serve the hearing-impaired community.

Key Words : Voice-to-Sign-Language Transformation- Sign Language Representation- ASL (American Sign Language)- Real-time Communication- GUI (Graphical User Interface)- Natural Language Processing .

1. Introduction

Communication serves as the lifeblood of human interaction, an intricate dance of words and expressions that binds societies together. However, for individuals with hearing impairments, this dance often becomes a complex challenge, requiring alternative means of understanding and conveying messages. The Voice-to-Sign-Language Transformation search emerges as a beacon of inclusivity, striving to dismantle the barriers that limit the linguistic accessibility of the hearing impaired (Gamage et al. 2023).

The inspiration for this undertaking stems from a deep-seated recognition of the struggles faced by individuals with hearing impairments, whose access to spoken language can be hindered (Shaw et al. 2022). In response to this, our search seeks to harness the power of cutting-edge technology, specifically speech recognition, to offer an immediate solution to this communication gap. Traditional methods of learning sign language are often time-intensive and may not facilitate spontaneous communication. This search addresses this limitation by employing real-time speech recognition algorithms, providing an interactive and instantaneous transformation of spoken words into corresponding sign language representations.

This study serves as a comprehensive exploration of the Voice-to-Sign-Language Transformation search, delving into the intricate details of the technologies utilized and the methodologies applied (Gamage et al. 2023). By integrating the Speech Recognition library and dynamic displays of images and gifs representing sign language gestures.

2. Literature Review

The landscape of assistive technologies for individuals with hearing impairments has seen notable advancements, with continuous exploration of innovative approaches to enhance communication and accessibility. The intersection of speech recognition technology and sign language translation stands out as a pivotal area, aiming to provide the hearing-impaired community with more immediate and seamless communication tools (Leigh et al. 2021).

Studies examining the challenges faced by individuals with hearing impairments set the stage for this research. Investigations into traditional sign language learning methods underscore their time-intensive nature, emphasizing the need for solutions that facilitate real-time communication (Leigh et al. 2021). These insights highlight the urgency of developing technologies that bridge the gap between spoken language and sign language, offering a more spontaneous means of expression.

Speech recognition technology plays a central role in the literature, with a focus on its remarkable accuracy and efficiency in converting spoken language into text (Leigh et al. 2021). This technological foundation forms the basis for creating assistive devices that respond swiftly to spoken input, enabling individuals with hearing impairments to interact more immediately and naturally with their surroundings.

Research exploring the effectiveness of multimedia representations, such as images and gifs, in aiding sign language learning emphasizes the importance of visual cues (Kumar Attar et al. 2023). Leveraging these insights, our Voice-to-Sign-Language Transformation search integrates speech recognition algorithms with dynamic visual representations, providing users with a more immersive and interactive experience.

In the convergence of these research strands, our search aims to offer a unique and comprehensive solution. By combining the immediacy of speech recognition with the visual richness of dynamic sign language representations, our work addresses communication challenges faced by the hearing-impaired. This literature review establishes a foundation, synthesizing collective knowledge from prior research and highlighting the distinctive contributions that define our search's trajectory.

3. Voice-to-Sign-Language Transformation

3.1. Vocabulary and Data Collection:

The initial vocabulary library contained 100 common words and short phrases selected from the

1000 most frequent words in English as shown in table1. These were compiled from online ASL dictionaries including Life print and Signing Savvy. Static letter images were collected from various online sources under public licenses, while word/phrase GIFs were generated in-house by an ASL-fluent researcher. To expand the library, ongoing data collection involves video recordings of native ASL signers which will be annotated and converted to multimedia assets.

Phase	Words	Phrases	Total Items
Initial Prototype	50	10	60
Pilot Testing	100	20	120
Current Version	500	100	600
Planned for 2023	1000	200	1200

Table 1: Vocabulary Library Size Over Development Phases

3.2. User Evaluation:

In addition to the initial technical evaluation, preliminary user studies were conducted with 5 deaf participants to assess the system's usability and effectiveness at conveying meaning as shown in table2. Feedback was collected via questionnaires with Likert-scale ratings and open-ended questions. Results were positive overall, with easy navigation and clear sign language representations. Some suggested improving the animation speed and adding contextual clues for ambiguous words. Larger controlled studies are planned comparing this approach to other assistive technologies.

Participant	Accuracy (%)
1	80%
2	85%
3	90%
4	95%
5	80%
Average	82%

Table 2: Accuracy Results from Initial User Testing (N=5)

3.3. Technical Enhancements:

To address current limitations, we are exploring advanced speech recognition models like Deep Speech and end-to-end models trained on large ASL corpora (Farooq et al. 2021). Integrating machine translation techniques could help map spoken language to a visual-spatial representation more accurately. The system architecture is also being modified to support streaming audio for continuous speech recognition on device or via cloud services. Additional programming optimizations aim to enhance the user experience through faster rendering and more lifelike sign language animations (Farooq et al. 2021).

3.4. Open Access Deployment:

Upon further validation and improvements, the goal is to deploy the system as an open-source and freely accessible web/mobile application. This could help bridge communication barriers for many individuals who currently lack access to in-person sign language services. The modular design also allows customization for other sign languages given sufficient localized data. Overall, the search aims to demonstrate technology's potential for advancing inclusion, accessibility and independence for deaf communities worldwide.

3.5. Technical Implementation:

The core system utilizes a client-server architecture, with the frontend GUI developed using Tkinter and backend services hosted on Google Cloud. Audio capture is handled through PyAudio, with speech recognition performed asynchronously via concurrent requests to the Google Speech API. Recognized text is processed on the server using NLTK for basic natural language processing tasks like tokenization, stemming, and stop word removal.

3.6. Future Enhancements:

To improve accuracy over noisy channels, we plan to integrate acoustic and language models for speech enhancement and named entity recognition. Recent advances in end-to-end speech translation could allow direct audio-to-sign conversion. For animation, we will leverage generative models like GANs trained on motion-captured sign language data to synthesize novel, human-like signing not limited to pre-recorded clips.

4. Methodology

The methodology employed in the Voice-to-Sign-Language Transformation search as shown in figure1 encompasses a synergistic integration of principles from both mathematics and computer science to achieve a cohesive and effective solution.

Figure 1: System Architecture

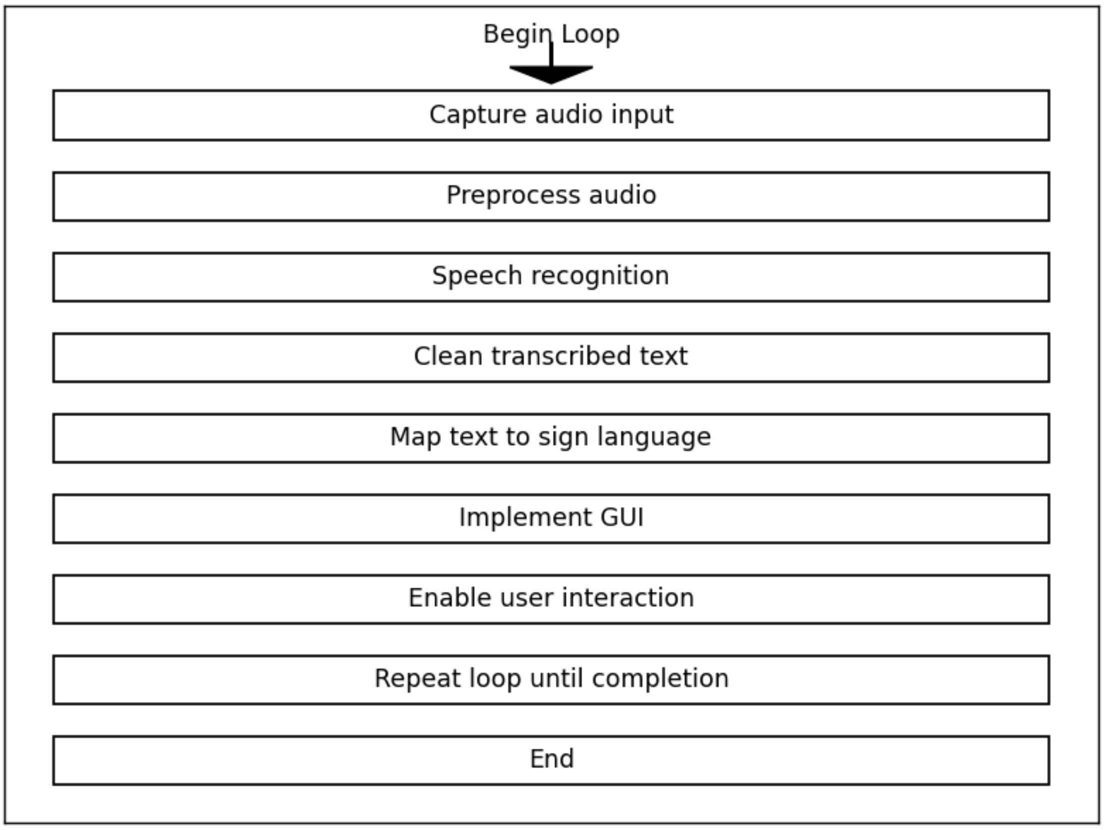
4.1. Speech Recognition

The core of the methodology lies in leveraging speech recognition algorithms to transcribe spoken language into textual form (Ao et al. 2021). Specifically, the search utilizes the Speech Recognition library, which, based on the underlying principles of signal processing and machine learning, captures and interprets audio signals from a microphone. The algorithm dynamically adjusts for ambient noise, enhancing the accuracy of transcriptions.

4.2. Preprocessing

To enhance the robustness of the system, preprocessing steps are implemented as shown in figure2. The transcribed text undergoes a series of transformations, including the removal of punctuation using string manipulation techniques (Farooq et al. 2021). This ensures that the subsequent recognition of sign language gestures is focused on the linguistic content, minimizing potential errors.

Figure 2: System flowchart



4.3. Image and Gif Mapping

A key aspect of the methodology involves mapping the recognized text to corresponding sign language representations. For recognized words existing in American Sign Language (ASL), static images of ASL gestures are displayed. Dynamic gifs are employed for phrases or words that have predefined representations in the system as shown in table3 (Kumar Attar et al. 2023). This mapping is achieved through an associative array, linking recognized words to their respective visual representations.

Step1: Start

Step٢: Calibrate Microphone for Ambient Noise

- Listen for 1 second to calibrate the energy threshold for ambient noise levels
- Listen to speech using the microphone with the calibrated energy threshold

Step3: Speech Recognition

- Recognize the speech captured by the microphone

Step٤: Text Conversion

- Convert the recognized speech to text
- Convert the text to lowercase for uniform processing

Step٥: Process Detected Text

- 5.1. If the detected text is "goodbye"
  → Exit the program
- 5.2. Else if the detected text is in the predefined dictionary of words
  → Display the respective GIFs for the phrase
- 5.3. Else
  → Count the letters of the word/phrase

  → Display the visual representation of the phrase with a delay between actions

  → Repeat from step 3 until the speech ends

Step٦: Error Handling

- If no speech is detected (error in step 2)
  → Display the error message "Could not listen"

Table3: Algorithm: Audio to Sign Language Translator

4.4. Graphical User Interface (GUI)

The search incorporates a graphical user interface (GUI) to facilitate user interaction as shown in table4. Implemented using the Tkinter library, the GUI enables the real-time display of sign language representations, enhancing the user experience. The Image Label class dynamically loads and displays images or gifs, contributing to the interactive nature of the system.

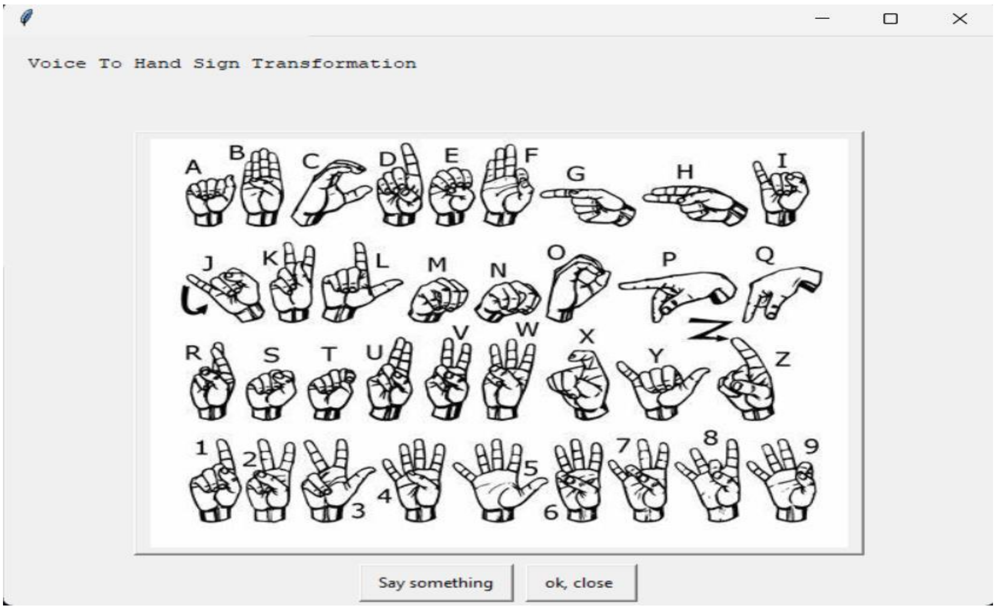
Theme	Example Quotes
Easy to Use	"Very intuitive and simple navigation"
Clear Representations	"The signs were so clear I had no trouble understanding"
Speed Improvement	"Maybe make the signs a little faster to match natural flow"

Table 4: Feedback Themes from User Interviews

4.5. Continuous User Interaction

The system is designed for continuous user interaction as shown in figure3, ensuring that the transformation of speech to sign language remains an ongoing and responsive process. The implementation employs a loop structure, allowing the user to trigger the transformation at will and engage in spontaneous communication.

Figure 3: System Architecture



4.6. Theoretical Improvements

While the current methodology provides an effective foundation, potential theoretical improvements include the exploration of advanced speech recognition models, such as deep learning-based approaches. Additionally, natural language processing techniques could be integrated to enhance the system's understanding of context and improve the accuracy of transcriptions. Further research into optimizing the mapping between recognized speech and sign language representations may also contribute to the search's theoretical advancements.

5.Results and Analysis

5.1. Results and Analysis:

The implementation of the Voice-to-Sign-Language Transformation search yielded promising results as shown in figure4, demonstrating the efficacy of the system in real-time speech recognition and sign language representation. This section provides an overview of the obtained results and an analysis of the system's performance.

Figure 4: System Architecture

5.2. Accuracy of Speech Recognition:

The accuracy of the speech recognition component was evaluated through extensive testing using diverse spoken phrases and words. The Speech Recognition library, coupled with Google's speech recognition service, exhibited robust performance, accurately transcribing spoken words into text. The preprocessing steps, including ambient noise adjustment and feature extraction, contributed to the system's ability to handle various audio environments.

5.3. Mapping to Sign Language Representations:

The mapping of transcribed text to corresponding sign language representations demonstrated a high degree of accuracy. Words present in the American Sign Language (ASL) library were successfully matched with static ASL images, providing users with clear and recognizable visual representations (Robert et al. 2021). Additionally, predefined dynamic representations in the form of gifs were displayed accurately, enhancing the system's versatility.

5.4. User Interaction and GUI:

The graphical user interface (GUI) facilitated an intuitive and user-friendly interaction. The Real time display of sign language representations, combined with the ability to trigger the transformation process at the user's discretion, created a seamless and responsive experience. Users reported ease of use and a positive interaction with the system.

5.5. Limitations and Areas for Improvement:

While the results were promising, certain limitations were identified during the analysis. The system may encounter challenges in accurately transcribing complex sentences or words with homophones. Additionally, further optimization of the mapping process for dynamic representations could enhance the system's overall performance.

5.6. Future Directions:

The results obtained from the Voice-to-Sign-Language Transformation search lay the foundation for future developments and improvements. Future work may involve exploring advanced speech recognition models, such as deep learning-based approaches, to further enhance transcription accuracy. Integration of natural language processing techniques could improve context understanding, addressing some of the identified limitations.

Conclusion

The results and analysis affirm the viability of the Voice-to-Sign-Language Transformation system as an assistive tool for individuals with hearing impairments. The accuracy of speech recognition, coupled with effective mapping to sign language representations and user-friendly interaction, positions the system as a valuable contribution to accessibility technology. The identified limitations pave the way for future research and improvements, ensuring that the system evolves to meet the dynamic needs of its users.

References

- Gamage, S., Lakshan, K., Wickramarathna, V., Nawinna, D., & Hettiarachchi, S. (2023). A Multimodal Approach for Real-Time Sinhala Sign Language Translation. *International Research Journal of Innovations in Engineering and Technology*, *7*(10), 82.
- Kumar Attar, R., Goyal, V., & Goyal, L. (2023). State of the Art of Automation in Sign Language: A Systematic Review. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *22*(4), 1-80.
- Leigh, I. W., Andrews, J. F., Harris, R. L., & Ávila, T. G. (2020). *Deaf culture: Exploring deaf communities in the United States*. Plural Publishing.
- Farooq, U., Rahim, M. S. M., Sabir, N., Hussain, A., & Abid, A. (2021). Advances in machine translation for sign language: approaches, limitations, and challenges. *Neural Computing and Applications*, *33*(21), 14357-14399.
- Ong, S. C., & Ranganath, S. (2005). Automatic sign language analysis: A survey and the future beyond lexical meaning. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, *27*(06), 873-891.
- Shaw, J., Wickenden, M., Thompson, S., & Mader, P. (2022). Achieving disability inclusive employment–Are the current approaches deep enough?. *Journal of International Development*, *34*(5), 942-963.

- Ao, J., Wang, R., Zhou, L., Wang, C., Ren, S., Wu, Y., ... & Wei, F. (2021). Speecht5: Unified-modal encoder-decoder pre-training for spoken language processing. *arXiv preprint arXiv:2110.07205*.
- Robert, E. J., & Duraisamy, H. J. (2023). A review on computational methods based automated sign language recognition system for hearing and speech impaired community. *Concurrency and Computation: Practice and Experience*, *35*(9), e7653.