**SIGN LANGUAGE TRANSLATOR WITH CHAT APP AND SPEECH SYNTHESIS USING COMPUTER VISION AND MACHINE LEARNING**

**BY**

**UDE, CHIDERAA EBENEZER**

**170408017**

**PROJECT IN THE DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING SUBMITTED TO THE FACULTY OF ENGINEERING IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF BACHELOR OF SCIENCE IN COMPUTER ENGINEERING**

**UNIVERSITY OF LAGOS**

**AKOKA, NIGERIA**

**DECEMBER 2023**

**CHAPTER ONE – INTRODUCTION**

**1.1 Background to the Study**

In today's world, clear and concise communication is paramount. For most people, this comes naturally through verbal interaction. However, for a significant portion of the population, this is not the case. According to the World Health Organization (WHO), more than 1.5 billion people (about one-fifth of the global population) live with hearing loss, of whom more than 400 million have disabling hearing loss [1]. While WHO estimates that up to 50% of such hearing loss can be prevented through public health measures [1], additional practical solutions are needed to circumvent the communication barriers it poses.

Sign language is a form of communication used by individuals with hearing or speech disabilities [2]. Alongside spoken languages, sign languages (SLs) constitute some of the most well-structured and organized communication systems globally. Similar to spoken languages, they vary according to the regions and communities from which they originate [3]. Some popular examples include American Sign Language (ASL), Arabic Sign Language (ArSL), Chinese Sign Language (CSL), and Italian Sign Language (LIS). These languages are primarily distinguished by their simultaneity, multimodality, and iconicity in the act of communication itself. Information is conveyed through hand shapes, movements, and oriented gestures [3].

Sign languages (SLs) have empowered many individuals with hearing or speaking impairments to connect with each other and participate in gatherings and conferences. This benefit for those with disabilities, however, has often required those without them to adapt their communication style—learning and using sign language, though unnatural for them. But what if a way existed for both parties—individuals with and without impairments—to interact freely in their preferred languages?

**1.2 Statement of the Problem**

The aim of this study is to investigate a solution that facilitates seamless communication between individuals who are hearing-impaired or speech-impaired and their non-impaired counterparts, allowing communication to flow in both directions. This is an area of significant interest because thorough exploration could spur further research in this field and open up avenues for investigating sign language as a reliable means of communication for people with such disabilities.

Schmalz [3] published a paper titled "Real-time Italian Sign Language Recognition with Deep Learning," which explored a method for processing hand gestures and movements from Italian Sign Language (LIS) into easily recognizable text using deep learning models. However, it focused solely on single-handed, isolated signs and couldn't handle double-handed or more complex ones. Chandrasekaran's Master's thesis [4] followed a similar approach but focused on American Sign Language. It presented a high-level overview of the problem but lacked a concrete implementation. The work in[2] examined a holistic solution, but it lacked a user-friendly interface for potential users to interact with.

This work aims to build a robust and practical solution for facilitating communication between individuals with hearing and speech disabilities and the general population. It will leverage deep learning models proposed by the aforementioned authors, integrating them into a user-friendly mobile application for both impaired and non-impaired individuals. This application will be bi-directional, enabling seamless two-way interlocution between both parties. This feature is highly relevant and in demand, as it has the potential to significantly increase participation in international conferences for individuals with these disabilities.

**1.3 Aim and Objectives of the Study**

The aim of this study is to develop and implement a real-time sign language to speech translator for integration into a mobile chat application. This translator, once designed, will be able to recognize a wide range of gestures and convert them into spoken language.

To achieve the previously mentioned goal, we have outlined the following specific objectives:

1. **Implementing a Robust Sign Language Recognition System**: To develop and deploy an efficient Sign Language (SL) recognition system capable of accurately interpreting diverse gestures.
2. **Integrating Real-Time Translation Mechanisms:** To incorporate real-time translation mechanisms using advanced computer vision techniques.
3. **Utilizing Machine Learning for Continuous Improvement and Adaptation:** To employ machine learning algorithms to continually enhance the system's adaptability and recognition accuracy.
4. **Ensuring User-Friendly Inclusive Design:** To design a user interface that considers the diverse needs of users, accommodating varying degrees of hearing and speech impairment.
5. **Testing and Validation with Diverse User Groups:** To conduct thorough testing and validation with users with diverse hearing abilities, speech patterns, cultural backgrounds, and sign language fluency and varying degrees of hearing and speech impairment.

**1.4 Methodology**

As previously discussed, the core of this project lies in implementing a bidirectional real-time sign language-to-speech translator. This will facilitate seamless communication between individuals with hearing and/or speech impairments and those without. This section outlines the systematic approach needed to achieve this goal. Each methodology will directly address one of the objectives previously mentioned.

Table 1: Project Methodology

|  |  |  |  |
| --- | --- | --- | --- |
| **S/N** | **Objectives** | **Material(s)** | **Method(s)** |
| 1. | Implementing a Robust Sign Language Recognition System | a. Visual Studio Code  b. Keras and TensorFlow | Convolutional Neural Networks (CNNs) |
| 2. | Integrating Real-Time Translation Mechanisms | a. Visual Studio Code  b. Python’s OpenCV libraries | Computer Vision (CV) |
| 3. | Utilizing Machine Learning for Continuous Improvement and Adaptation | a. Visual Studio Code  b. Python’s machine learning libraries | Machine learning algorithms for continuous improvement |
| 4. | Ensuring User-Friendly Inclusive Design | a. Visual Studio Code  b. Terminal emulator  c. Web browser | Responsive frontend website development |
| 5. | Testing and Validation with Diverse User Groups | Individuals with hearing and speech impairments from the university community | Popular sampling techniques |

**1.5 Scope of the Study**

This project specifically aims to implement a user-friendly and highly responsive interface for potential end users to interact with the sign language translator in real time.

**1.6 Significance of the Study**

This study aims to fuel advancements in sign language translation. If successful, it is expected to spur significant investment from tech entrepreneurs and business owners, thereby contributing to global economic growth and raising the living standards of people around the world.

**1.7 Research Questions/Hypothesis**

Can a deep learning model (such as Convolution Neural Networks) be employed in American Sign Language recognition and translation and eventually augmented into a mobile application?

**1.8 Outline of the Project Report**

This report is structured according to the outline given below.

**Chapter 1: Introduction**

This chapter opens by discussing how sign language has empowered the estimated billion individuals living with hearing and/or speech disabilities. It then introduces the research problem: designing and implementing a sign language-to-speech translator to facilitate communication between people with these impairments and those without. Subsequently, the chapter specifies the study's aims, objectives, methodologies, scope, and significance. Finally, it outlines the structure of the entire project report.

**Chapter 2: Literature Review**

This chapter presents a comprehensive review of past research in the field of sign language to speech translation. Readers are introduced to the key advancements made to date, highlighting the methodologies, tools, and datasets employed by each researcher.

**Chapter 3: Methodology**

Here, the methods for building the sign language translator are discussed in detail. The author first introduces several models and then narrows down the selection to a more robust and efficient one: a Convolutional Neural Network (CNN).

**Chapter 4: Results and Discussion**

In this chapter, the author presents the results of the deep learning model used to implement the sign language translator. The model's performance across multiple datasets, as well as its integration into the chat app, are thoroughly discussed.

**Chapter 5: Conclusion and Recommendation**

This chapter summarises the work that was done. Conclusions are drawn concerning the reliability of the chat application and its integration into popular social media and instant messaging platforms. Recommendations for further study are also made here.

**References**

All literature consulted or directly cited in this study is listed here.

**Appendices**

Snapshots of the chat app, as well as the source code for the sign language translator, are provided here.

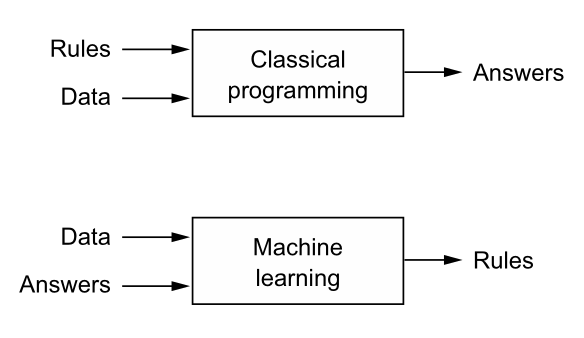
**CHAPTER TWO – LITERATURE REVIEW**

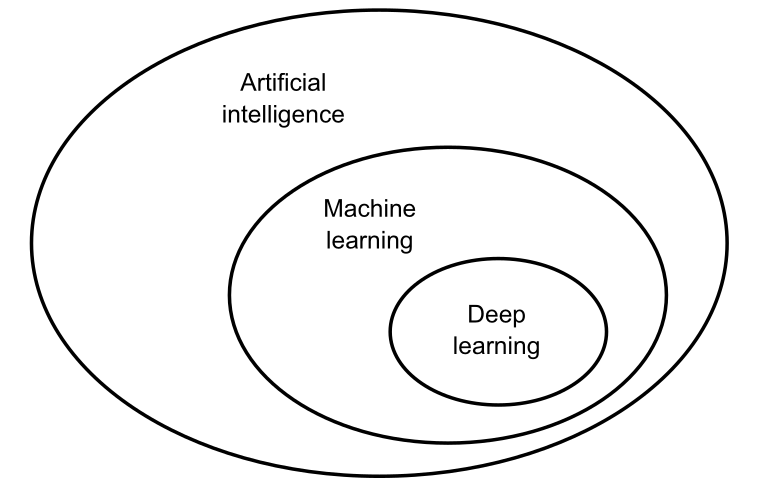
**2.1 American Sign Language (ASL) Translation**

American Sign Language is one of the many sign language conventions used worldwide. It is a manual-visual language, produced by the hands and perceived by the eyes [5]. For people with hearing and/or speech impairments, ASL is an accessible and natural form of communication, often their only reliable means. For non-impaired individuals, however, understanding and conversing in ASL can be challenging, as they primarily rely on spoken or written language. This creates a communication barrier between the two groups. Therefore, the need for a translator arises to bridge this gap and effectively serve both sides. This is precisely the aim of this paper. The author proposes a real-time sign language to speech translator specifically based on the American Sign Language system (ASL).

**2.2 Deep Learning Models**

The translator thus intended must necessarily be some form of computer program, as it would be tedious for a human to perform the translation continuously. However, this program would be unlike classical programming, where a set of rules and data are fed into a model to generate some output (see Figure XYZ). This is because there are no explicit algorithms defining the mapping of hand gestures and signs to meaningful speech and/or text. The only possible way to make this translation is to learn from data that has, in some way not privy to the public, been identified with certain words. Thus, learning is being achieved—and this is, in fact, what machine learning is all about. It looks at input data and the corresponding output and infers what the rules producing it should be. A machine learning system is trained, rather than explicitly programmed. [6].

Figure 2: Machine learning: a new programming paradigm [6]

Figure 1: Artificial intelligence, machine learning and deep learning [6]

Deep learning distinguishes itself in the landscape of machine learning by progressively extracting increasingly meaningful representations from data across multiple layers. Each layer builds upon its predecessor, culminating in modern models with hundreds of levels. This iterative process, the depth, differentiates deep learning from other, shallower approaches that typically employ one or two layers. To achieve this multi-layered learning, deep learning leverages stacked artificial neural networks. While inspired by neurobiology, these networks are not direct replicas of the brain. Deep learning stands on its own as a powerful mathematical framework for extracting rich data representations, leaving behind comparisons to neuroscience [6]. In this work, Convolutional Neural Networks (CNNs) will be employed to achieve sign language translation.

**2.3 Related Work**

Several research areas within sign language translation have benefited from machine learning. For instance:

Kumar, *et al.* in [7] developed EasyTalk, a sign language translator facilitating bi-directional communication between individuals in Sri Lanka with speech and/or hearing impairments and those without. It comprises four key components enabling its overall operation. First, the Hand Gesture Detector identifies hand signs using pre-trained models. Next, the Image Classifier analyzes the captured signs and translates them into text. Then, the Text and Voice Generator outputs the translated information in either text or audio format. Finally, the Text-to-Sign Converter transforms written English back into animated sign language gestures.

Abiyev, *et al.* in [8] utilized object detection and classification to achieve sign language translation with deep convolutional neural networks. The system they designed employed Single Shot MultiBox Detection (SSD) to locate hand gestures. They also introduced a novel hybrid structure that fused the Inception v3 architecture for feature extraction with a Support Vector Machine (SVM) classifier. Their model was built upon a sign language fingerspelling dataset and enabled efficient sign language translation from detected gestures. Their results demonstrated the effectiveness of this hybrid approach in sign language translation.

Naranjo-Zeledón, *et al.* in [9] propose a comprehensive and systematic mapping of technologies that facilitate sign language translation. Their work adheres to best practices in software engineering, drawing on systematic reviews and emphasizing the accessibility of human-computer interaction, natural language processing, and education – all areas directly relevant to software engineering within the Association for Computing Machinery (ACM) computing classification system.

Driven by a pressing need for inclusive communication in South Africa, Madahana, *et al.’s* [10], study aimed to bridge the gap between cutting-edge real-time speech-to-sign language translation technology and its limited application for the hearing-impaired on the African continent. Their research highlighted a dearth of evidence surrounding the use of AI and machine learning in such solutions. To address this gap, the authors proposed an AI-powered, real-time South African translator specifically designed to facilitate seamless two-way communication between individuals with and without hearing disabilities. This solution harnessed the potential of AI to translate spoken South African languages into accurate and nuanced sign language in real-time, empowering the hearing-impaired community and fostering greater social inclusion.

Chandrasekaran in [4] addressed sign language recognition and translation using the American Sign Language Alphabet dataset. The author selected a Convolutional Neural Network (CNN) for both tasks, as the data consisted of images. The resulting model predicted the signs, displaying them in text and converting them to audio using Google Text-to-Speech. The model achieved 94% accuracy after 10 epochs. However, it exhibited overfitting, which future work should address while aiming for further improvement.

Bantupalli and Xie proposed a novel application in [2] aaimed at bridging the communication gap between signers and non-signers through sign-language-to-text translation. This video-based system extracted both temporal and spatial features from video sequences. It recognized spatial features by leveraging a convolutional neural network (CNN) architecture known as Inception. Subsequent temporal features were captured by training a recurrent neural network (RNN) upon them. Training data for this endeavor was sourced from the American Sign Language Dataset.

The work conducted by the aforementioned authors was geared primarily toward building the sign-language-to-speech and sign-language-to-text solution. However, there was no practical implementation for potential end users. Thus, as previously highlighted, the goal of this work is to fill that gap. A robust and fully functioning mobile application containing the sign-language-to-speech solution is the expected outcome.

**2.4 Future Work**

While this work focused primarily on building a sign-language-to-speech translator using American Sign Language (ASL), it also explored the possibility of implementing the same technology for other sign language conventions. These include Arabic Sign Language (ArSL), Indian Sign Language (ISL), Sri Lankan Sign Language, Chinese Sign Language (CSL), Italian Sign Language (LIS), and many more.

Further effort should be directed towards expanding the solution beyond a mobile app, encompassing web and desktop applications. Integrating the sign-language-to-speech translator seamlessly into popular social media and messaging platforms like WhatsApp, Facebook, Instagram, and others would significantly enhance accessibility and connectivity.

**2.5 Summary**

This paper aims to develop a real-time American Sign Language (ASL) to speech translator using machine learning, particularly deep learning via Convolutional Neural Networks (CNNs). Traditional programming with explicit rules is unsuitable for sign language translation due to its complex grammar and lack of defined sign-to-speech mappings. Therefore, the translator learns from data, similar to how humans do. Previous research in sign language translation with machine learning is reviewed, highlighting projects like EasyTalk, hybrid deep learning approaches, and systematic mapping frameworks. These demonstrate the feasibility of the proposed approach. This work's unique contribution is an easy-to-use mobile application for real-time ASL-to-speech translation, bridging the gap between existing research and practical use.

**REFERENCES**

[1] ‘Deafness and hearing loss’. Accessed: Dec. 29, 2023. [Online]. Available: https://www.who.int/health-topics/hearing-loss

[2] K. Bantupalli and Y. Xie, ‘American Sign Language Recognition using Deep Learning and Computer Vision’, in *2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA: IEEE, Dec. 2018, pp. 4896–4899. doi: 10.1109/BigData.2018.8622141.

[3] V. J. Schmalz, ‘Real-time Italian Sign Language Recognition with Deep Learning’.

[4] S. Chandrasekaran, ‘American Sign Language Recognition and Translation using Deep Learning and Computer Vision’.

[5] S. K. Liddell, *American Sign Language Syntax*. Walter de Gruyter GmbH & Co KG, 2021.

[6] F. Chollet, *Deep learning with Python*, Second edition. Shelter Island: Manning Publications, 2021.

[7] D. Manoj Kumar, K. Bavanraj, S. Thavananthan, G. M. A. S. Bastiansz, S. M. B. Harshanath, and J. Alosious, ‘EasyTalk: A Translator for Sri Lankan Sign Language using Machine Learning and Artificial Intelligence’, in *2020 2nd International Conference on Advancements in Computing (ICAC)*, Malabe, Sri Lanka: IEEE, Dec. 2020, pp. 506–511. doi: 10.1109/ICAC51239.2020.9357154.

[8] ‘Sign Language Translation Using Deep Convolutional Neural Networks’, *KSII Trans. Internet Inf. Syst.*, vol. 14, no. 2, Feb. 2020, doi: 10.3837/tiis.2020.02.009.

[9] L. Naranjo-Zeledón, J. Peral, A. Ferrández, and M. Chacón-Rivas, ‘A Systematic Mapping of Translation-Enabling Technologies for Sign Languages’, *Electronics*, vol. 8, no. 9, p. 1047, Sep. 2019, doi: 10.3390/electronics8091047.

[10] M. C. Madahana, K. Khoza-Shangase, N. Moroe, D. Mayombo, O. Nyandoro, and J. Ekoru, ‘A proposed artificial intelligence-based real-time speech-to-text to sign language translator for South African official languages for the COVID-19 era and beyond: In pursuit of solutions for the hearing impaired’, *S. Afr. J. Commun. Disord.*, vol. 69, no. 2, Aug. 2022, doi: 10.4102/sajcd.v69i2.915.

[11] S. Chandrasekaran, ‘American Sign Language Recognition and Translation using Deep Learning and Computer Vision’.