Transcending Barriers: A Comprehensive Framework for Sign Language Video Translation and Speech Generation into Low-Resource Languages (LRLs) in Nigeria

IPADEOLA LADIPO

INTRODUCTION

Effective communication is a fundamental aspect of human interaction, fostering connections and understanding within societies. However, individuals in the Nigerian Sign Language (NSL) community often encounter challenges when communicating with those outside their immediate circles. Sign language, a crucial means of interpersonal communication for the hearing-impaired, faces limitations in bridging the gap with the broader society.

According to the World Health Organization (WHO), an estimated 5.3% of the global population, around 430 million people, experience hearing disabilities, with a pronounced prevalence in sub-Saharan Africa (Kolawole, et al. 2022). Despite this considerable prevalence, developing countries lack adequate communication support facilities for the hearing-impaired, resulting in children with hearing loss being deprived of formal schooling, leading to higher unemployment rates, and contributing to social isolation and stigmatization. In light of these challenges, advancements in Computer Vision, Natural Language Processing (NLP), Generative AI and artificial intelligence in general, have presented innovative solutions to address hearing disabilities. However, the scarcity of sign language data in these regions and the complexity of deploying advanced tools in real-life environments hinder the widespread adoption of such solutions (Kolawole, et al. 2022).

This project focuses on the development of an Automatic Sign Language Translation system, specifically targeting the translation of Nigerian Sign Language into one of the Low-Resource Languages (LRLs) in Nigeria. The significance of this project lies in its potential to enhance inclusivity and create employment opportunities for the NSL community, particularly in the education sector and the workforce sector of Nigeria.

PROBLEM DESCRIPTION

There are about 70 million deaf people who use sign language as their first language or mother tongue (Mulwafu, Kuper and Ensink 2016). It is also the first language and mother tongue of many hearing people and some deaf-blind people (tactile sign languages). Each country has one or sometimes two or more sign languages, although different sign languages can share the same linguistic roots in the same way as spoken languages do. Sign language (also signed language or simply signing) is a language which uses manual communication and body language to convey meaning, as opposed to acoustically conveyed sound patterns.

This can involve simultaneously combining hand shapes, orientation and movement of the hands, arms or body, and facial expressions to fluidly express a speaker's thoughts.

In the vibrant and diverse community of Nigeria, the Sign Language community encounters formidable challenges that impede their full integration into society. There are several issues faced by the NSL community in Nigeria.

Educational Challenges: One of the primary obstacles faced by the NSL community in Nigeria is the limited access to quality education. Scarce specialized resources, a shortage of trained educators, and a dearth of appropriate learning materials contribute to an educational landscape that often fails to cater to the unique needs of individuals with hearing impairments. The absence of a widespread understanding and promotion of inclusive education exacerbates this problem.

Communication Barriers: Communication lies at the heart of human interaction, yet for the NSL community in Nigeria, this fundamental aspect is fraught with challenges. The percentage of the English-speaking population in Nigeria is about 53-60%, with the remaining divided among the native languages (Yoruba, Igbo, Hausa, etc.) (Sasu 2021). The prevalent use of spoken languages without sufficient support for sign languages, such as the Nigerian Sign Language (NSL), erects communication barriers. A lack of awareness and training in sign language among the general population further perpetuates the isolation of the NSL community. In fact, for one, if a parent/loved one of a deaf and hard-of-hearing individual wants to learn Sign Language in order to enable communication, he or she would have to do deep research as schools that teach sign language are very rarely even in developed cities, like Lagos and Abuja. The learning curve for hearing individuals is high, as they find it very difficult to quickly adapt and learn. Learning usually takes up to a year or two before one can be able to sign just phrases. In rural areas, the schools, awareness and opportunities to learn Sign language are close to nil, as NSL individuals usually communicate using gestures and non-understandable sounds, which is a serious communication challenge.

Limited Employment Opportunities: The job market in Nigeria poses another formidable challenge for the NSL community. Often, employers lack awareness of the capabilities of individuals with hearing impairments, leading to discrimination and restricted employment opportunities. The absence of accommodations in the workplace further hinders the professional advancement of this community (Dere, et al. 2023).

Healthcare Accessibility: Access to healthcare services is a critical concern for the NSL community. Communication barriers with medical professionals, due to a lack of interpreters or accessible communication tools, result in inadequate healthcare. This issue, when combined with the general shortcomings in the healthcare system, compounds the difficulties faced by individuals with hearing impairments. The prohibitive cost of hearing aids and other assistive devices further exacerbates this technological divide.

Social Stigma and Discrimination: Deep-seated societal misconceptions and stigmas surrounding hearing impairments contribute to discrimination and exclusion. Limited awareness campaigns and community engagement perpetuate negative attitudes toward the NSL community, hindering their social integration and well-being.

Utilizing technological advances in the Artificial intelligence field, a machine learning system can be developed that makes the communication barrier less effective and helps in reducing/resolving the challenges previously highlighted.

PROPOSED METHOD

Developing an NSL Translation system, that translates sign language into spoken words, one of the native languages (Yoruba), entails three fields of Artificial intelligence. They are Computer Vision, Natural Language Processing (NLP), and Generative AI. As such, the proposed method for developing this solution as be broken down into those three sections.

Computer Vision: In this dedicated section, the primary objective is the development of a sophisticated model capable of real-time ingestion of Nigeria Sign Language (NSL) videos and subsequent translation of the conveyed signs into text. Leveraging the well-established prowess of Transformer models in addressing sequence-to-sequence challenges, the process will be initiated by implementing, training, and thoroughly analyzing a baseline transformer model. This foundational step will serve as a solid benchmark for subsequent enhancements. Following the establishment of our baseline model, a comprehensive exploration awaits. Delving into a myriad of advanced models, cutting-edge techniques, and state-of-the-art architectural innovations to further elevate the performance of the system. This iterative approach ensures a thorough investigation into the most effective methodologies available in the current landscape of machine learning and artificial intelligence.

To fuel the exploration and validation, a diverse and robust set of datasets will be used. The primary contributors to our model's training and evaluation will be the Google American Sign Language (ASL) dataset (Google 2023), renowned for its richness and authenticity, and the How2Sign dataset (Duarte, Palaskar and Ventura 2020), a valuable resource in its own right. Nigerian Sign Language uses American Sign Language(ASL) as its foundation with only a few changes (Kolawole, et al. 2022).These datasets collectively provide a broad spectrum of ASL expressions, enabling our model to learn and generalize effectively across various signing styles and contexts. Through this meticulous selection of datasets, the aim is to fortify the model's adaptability and efficacy in real-world scenarios.

Natural Language Processing (NLP): Within this dedicated section, our focus is on the translation of English text into a low-language resource, specifically Yoruba. To achieve this transformation, this is done by harnessing the capabilities of Meta's cutting-edge solution—the No Language Left Behind pre-trained model(Team, NLLB 2022). This model stands as a testament to the commitment to inclusivity in language translation, being specifically designed to enhance the accuracy and effectiveness of translations for low-resource languages. The No Language Left Behind model, having undergone meticulous pre-training, emerges as a powerful tool tailored to the nuances of low-resource linguistic contexts.

Generative AI: Our next critical phase involves the creation of a generative model tasked with transforming the translated text into spoken words. This intricate process necessitates the development and rigorous training of a specialized model designed to eloquently articulate the linguistic nuances captured in our translated output. This endeavour represents a crucial bridge between the realm of text-based communication and the auditory landscape, paving the way for a more immersive and inclusive user experience. To enrich our generative

model, we will harness the capabilities of a pre-existing powerhouse—the Massively Multilingual Speech (MMS) model curated by Meta. This pre-trained model, renowned for its multilingual proficiency, serves as a formidable foundation for our endeavours. By integrating this MMS model into our framework, we unlock a vast reservoir of linguistic knowledge, laying a robust groundwork for the generation of diverse and authentic spoken expressions. The synergy between our newly developed generative model and the pre-trained MMS model is poised to elevate our project to unprecedented levels of sophistication.

To fine-tune the MMS model for the intricacies of our task, we will draw upon the wealth of the IròyìnSpeech dataset (Ogunremi, et al. 2023). This dataset, carefully selected for its relevance and diversity, provides a nuanced collection of speech data that aligns seamlessly with our objectives. Through this meticulous fine-tuning process, we ensure that our generative model not only meets but exceeds expectations in delivering articulate and contextually rich spoken output.

Moreover, the commitment to enhancing each section extends to the implementation of a diverse array of techniques, optimization strategies, and vigilant monitoring skills. These initiatives are poised to elevate the performance and efficiency of each facet of the project. Here's an overview of the methodologies to be employed:

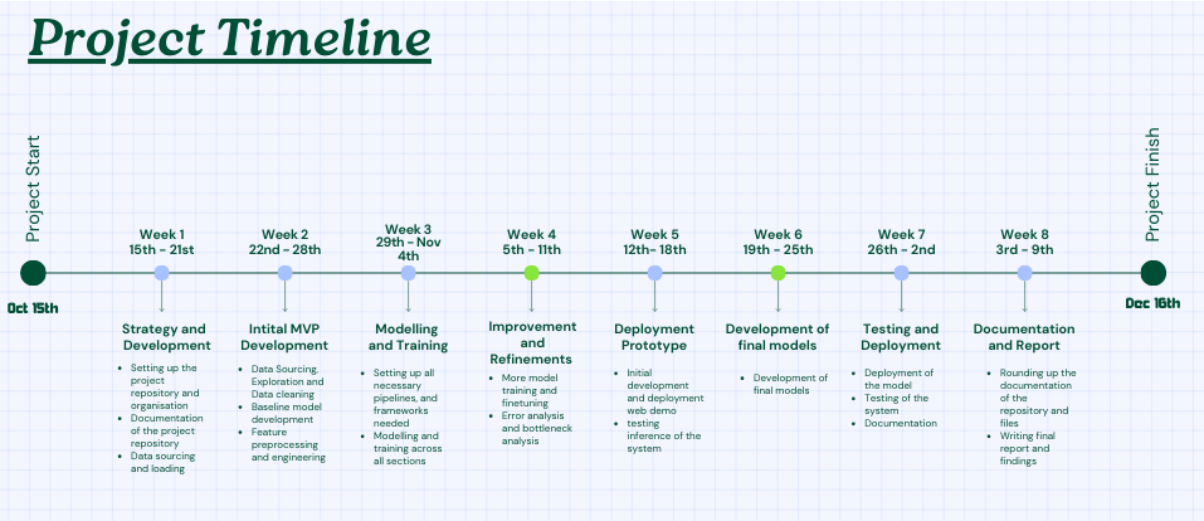
1. **Distributed Training with GPU Acceleration:** Embracing the power of parallel processing, leverage GPU resources for distributed training. This strategic move aims to significantly reduce training time, allowing to expedite the model development process and achieve faster convergence.
2. **Monitoring with Weights and Biases Framework and TensorBoard:** To maintain a vigilant eye on the health and performance of the models, implementation a robust monitoring system. The Weights and Biases framework, known for its comprehensive tracking capabilities, and TensorBoard, a powerful visualization tool, will be instrumental in providing insights into training metrics, model behaviour, and potential bottlenecks.
3. **Deployment via FastAPI**
4. **Error Analysis**

Proposed System Design Diagram

PROPOSED TIMELINE

Using the phrase, **“Fail fast, iterate and improve.”,** the proposed timeline is set up to be able to develop a working prototype as soon as possible and then try to improve its efficiency and accuracy.

Given that, the proposed timeline is drawn below:



CONCLUSION

References

Adeyanju, A, O.O Bello, and M.A Azeez. 2022. “Development of an American Sign Language Recognition System using Canny Edge and Histogram of Oriented Gradient.” *Nigeria Journal of Technological Development.*

Dere, Mustapha Deji, Oluwabukola Rashidat Dere, Adewale Adesina, and Aliyu Rufai Yauri. 2023. “AN end to end fraweork for translation of American Sign Language to low resource languages in Nigeria.” *Scientific African.*

Duarte, A, S Palaskar, and L Ventura. 2020. “How2Sign: A Large-scale Multimodal Dataset for Continuous American Sign Language.”

Google. 2023. *Google - American Sign Language Fingerspelling Recognition.* Accessed October 14, 2023. https://www.kaggle.com/competitions/asl-fingerspelling/data.

Kolawole, Steven, Opeyemi Osakuade, Nayan Saxena, and Babatunde Olorisade. 2022. “Sign-to-Speech Model for Sign Languae Understanding: A Case Study of Nigeria Sign Language.” *International Joint Conference on Artificail Intelligence.*

META. 2022. “No Language Left Behind: Scaling Human-Centered Machine Translation.”

Mulwafu, W, H Kuper, and R Ensink. 2016. “Prevalence and causes of hearing impairment in Africa.”

Ogunremi, Tolulope, Kola Tubosun, Anuoluwapo Aremu, Iroro \Orife, and David Ifeoluwa Adelani. 2023. “ÌròyìnSpeech: A multi-purpose Yorùbá Speech Corpus.”

Sasu, Dokua Doris. 2021. *Languages spoken in Nigeria as of 2021, by number of speakers(in 1,000s).* 21 January. Accessed October 12, 2023. https://web.archive.org/web/20221101042810/https://www.statista.com/statistics/1285383/population-in-nigeria-by-languages-spoken/#:~:text=Following%20this%20were%20Yoruba%2C%20Nigerian,Nigeria%2C%20with%2060%20million%20speakers.

Team, NLLB, Marta R., James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan et al. "No Language Left Behind: Scaling Human-Centered Machine Translation." *ArXiv*, (2022). Accessed October 17, 2023. /abs/2207.04672.