# AI-Driven Sign Language Interpretation for Nigerian Children at Home

**Ifeoma Nwogu**[1,2] , **Roshan Peiris**[2] , **Karthik Dantu**[1] , **Ruchi Gamta**[2] and **Emma Asonye** [3]

[1]Dept. of Computer Science and Engineering, University at Buffalo, NY
[2]Golisano College of Computing and Information Sciences, Rochester Institute of Technlogy, NY
[3]Dept. of Africana Studies, University of New Mexico, NM
inwogu@buffalo.edu, rxpics@rit.edu, kdantu@buffalo.edu, rxgntm@rit.edu, easonye@unm.edu

## Abstract

As many as three million school age children between the ages of 5 and 14 years, live with severe to profound hearing loss in Nigeria. Many of these Deaf or Hard of Hearing (DHH) children developed their hearing loss later in life, non-congenitally, hence their parents are hearing. While their teachers in school often readily and effectively communicate with them in "dialects" of American Sign Language (ASL), the unofficial sign lingua franca in Nigeria, communication at home with other family members is challenging and sometimes non-existent. This results in adverse social consequences including stigmatization, for the students.

With the recent successes of AI in natural language understanding, the goal of automated sign language understanding is becoming more realistic, using neural deep learning technologies. To this effect, the proposed project aims at co-designing and developing an ongoing AI-driven two-way sign language interpretation tool that can be deployed in homes, to improve language accessibility and communication between the DHH students and other family members. This ensures inclusive and equitable social interactions which can promote life-long learning opportunities for the students outside of the school environment.

## 1 Introduction

Deaf and Hard of Hearing (DHH) individuals constitute a significant portion of the population in Nigeria, where as many as 22 million Nigerians, in an estimated population of about 155 million people suffer from severe to profound hearing loss and are considered as *persons with disabilities* [Eleweke, 2002; Treat, 2016]. Studies have also indicated that up to 14% of school age children in Nigeria between the ages of 5 and 14 suffer from such severe hearing loss [McPherson and Swart, 1997]. These age ranges make up some of the largest population groups in Nigeria. In addition, as many as 84% of the deaf population in Nigeria remain under-educated and economically underdeveloped [Eleweke *et al.*, 2015], indicating a close connection between deafness and poverty in the country [Asonye, 2016].

In an extensive study of the Deaf culture in Nigeria [Asonye *et al.*, 2018], the authors collected data from a total of six primary and secondary schools for the Deaf, from the Northcentral and Southeastern regions of Nigeria which are traditionally very different cultural settings. They found the Deaf cultures across the regions to be similar, having comparable socioeconomic and assimilation challenges. For example, the earliest age that DHH children started kindergarten was about 6-7 years, the age where their hearing peers were already in the equivalent of 2nd or 3rd grade in elementary school. Similarly, students could be as old as 15 years, when starting elementary school, significantly older age ranges than their hearing peers.

As reported by the Nigerian Deaf study [Asonye *et al.*, 2018], about 75% of pupils/students observed in their study did not have congenital hearing loss, but rather developed deafness because of various diseases contacted in childhood[1]. Many parents of the deaf students therefore are themselves not deaf and for many cultural reasons, do not embrace the Deaf culture.

Early language acquisition has been shown to involve linguistic and neural processes that have to do with the post-natal brain development interacting with its environment [Grimshaw *et al.*, 1998], In the absence of social environments that can provide the DHH children with opportunities for communicative experiences, early language acquisition is missed and the associated neural, cognitive and linguistic processes are delayed or missed [Hoff, 2006]. In a 2018 study on the quality of life of deaf students in Southwest Nigeria [Jaiyeola and Adeyemo, 2018], it was found that stigmatization of deafness originated in the students' families and propagated to their immediate communities primarily due to problems with communication. The study stated that this communication gap between the Deaf students and their hearing family members continued into the years, sometimes having adverse consequences; the children are often discriminated against and excluded, they experience inequalities in society, and are exposed to vulnerabilities that tend to leave them behind.

[1]More details of the various causes of hearing loss they observed can be found in the Nigerian Deaf study [Asonye *et al.*, 2018].

AI research on automating two-way Sign Language (SL) translation can play a critical role in bridging this communication and engagement gap for Deaf students at home. This is a very challenging AI problem but one that can allow signers to provide and receive information in their own natural language, without requiring their interlocutors to learn a new language.

### 1.1 Problem Statement

*We propose an iterative end-to-end approach to deploying AI-driven sign language translation devices based on existing sign language tutoring processes. Such devices can provide 2-way sign language interpretation services to young DHH students and their families in a pre-identified study area in Nigeria, to foster inclusivity at home, among hearing family members and peers.*

### 1.2 Link to Specific SDGs and LNOB

This project goes beyond the traditional research objectives by connecting to specific *UN Sustainable Goals - "Quality Education", "Decent Work and Economic Growth", and "Reduced Inequalities"*. Improving language accessibility and communication tools for DHH students in their home settings ensures inclusive and equitable social interactions and promotes lifelong learning opportunities for them outside of the school environment. Ensuring this creates increased opportunities for full and productive future employment for the DHH children, thus resulting in reduced inequalities.

Similarly, the **Leave no one behind (LNOB)** principle aims to combat discrimination and rising inequalities within and amongst countries. The UN 2030 agenda [UN Department of Economic and Social Affairs, 2023] cites that a major cause of people being left behind is persistent forms of discrimination,, resulting in individuals, families and whole communities such as the Deaf community in developing countries such as in Nigeria, being marginalized and excluded [Chapple, 2019; Leigh *et al.*, 1998; Best, 2015]. Providing easier methods of communication between DHH students and their hearing family members and peers can potentially reduce discrimination, exclusion and unconscious biases; it can foster acceptance within the community and also diminish the vulnerabilities that leave DHH children behind, undermining their individual potentials.

To this end, this proposed project is in collaboration with Save the Deaf and Endangered Languages Initiative (S-DELI)[2], a community based nongovernmental organization (NGO) committed to the development of Deaf children and adults in Nigeria. S-DELI is comprised of (i) a research component that documents collects data about sign languages and Deaf demographic data across Nigerian Deaf communities; (ii) an advocacy and outreach component that successfully pushed the policy to include the implementation of Newborn Hearing Screening and Early Intervention Program (NBHS/EI) in Nigeria; and (iii) an interpretation service component that trains and provides interpreting services for deaf individuals and groups, families and corporate bodies.

[2]https://www.s-deli.org/

The proposed project is also in close collaboration with the Demonstration School for Deaf Children (DSDC), located in Kaduna, in Northern Nigeria. DSDC was founded by Brenda Woosman, a Canadian citizen in 1987, as she visited Northern Nigeria where she learned about the plight of deaf children in the area. Currently operating as non-profit organization, DSDC is partially supported by the Kaduna State Ministry of Human Services and Social Development (KDHSSD), a governmental organization focused on developing the rights and privileges of women, children, the socially disadvantaged and physically challenged in the State. There are currently 115 students from kindergarten 1 through Primary six all the way to Junior Secondary school level 3. The school has 13 teachers of which 7 are proficient in sign language. The Principal of DSDC, Mrs Victoria Adesina is a Project Collaborator.

Lastly, we are also collaborating with the National Technical Institute for the Deaf (NTID) via Ruchi Gamta, a faculty member and Instructor. NTID is the first and largest technological college in the world for DHH students [McCarthy, 2018] and is one of the nine colleges at the Rochester Institute of Technology (RIT), where three of the Investigators on this proposal are either current or past faculty members. More on this will be discussed in Section 4.

### 1.3 A Brief History of ASL in Nigeria

American Sign Language (ASL) is the primary sign language used in Nigeria, although over the years, various "dialects" of it are used across different regions of the country. ASL was introduced to the Nigerian Deaf Community by Andrew Forster, a Black American missionary and Galladuet University alumnus. Foster established the Ibadan Mission School for the Deaf in Southwestern Nigeria in 1960, and this school is credited to have significantly impacted the development and advancement of formal deaf education across Nigeria.

Till date, ASL along with its regional variants, is the primary sign language used in the schools for the Deaf, special education centers and deaf units established within regular schools across the country. *But in addition to ASL being the "sign lingua franca" in Nigeria, there are many different native sign languages that we will be interested in engaging with and documenting over the course of this work.*

## 2 Strategy and Case Study

The project will comprise of four major research thrusts: (i) user-centered interface design; (ii) 2-way neural ASL translation development (detailed description in Section 3; (iii) timely realization of the interpretation device; and (iv) field evaluation study. The project will be implemented across two study areas, in the US and in Nigeria where the device will be deployed to parents and family members, to enhance their communication with the DHH students at home.

**User-centered Interface Design:**
This component of the research will include iterative co-design sessions engaging various stakeholders such as DHH students, hearing students, Instructors, and Interpreters with the goal of developing a working 2-way sign-language interpreting agent. *It is important to note that the agent is not a replacement for human interpreters*, rather it is can act as

an agent to bridge the hearing-signing communication gap in cases when human interpreters are not readily available. The initial design work will be carried out at RIT. .

Because of the interdependence nature of the project, this thrust will employ a deliverable-oriented approach [Hakiel, 1995] to the user-centered design, so that the outputs of the design can guide the development of other research products in the project pipeline.

### Timely Realization of the Interpretation Device

Previous research has established the importance of timing [Bai *et al.*, 2022] and pauses in the use of animations [Alkhazraji *et al.*, 2018] for ASL among the Deaf population. In this research thrust, we seek to answer two primary research questions in realizing ASL signing using robots: **RQ1**: How can we autonomously detect, track and follow the intended communicator for two-way ASL interpretation? and **RQ2**: How can we ensure end-to-end timeliness of communication in accordance to accepted ASL signing between humans to best use an avatar to be an ASL interpreter? **RQ3**: How can we constrain the 2-way translation AI engine sufficiently, to successfully operate on a small mobile device that can be deployed in a low-resource third-world country?

We aim to address these research questions in this thrust.

### Field Evaluation Study

The project will be initially iteratively designed, developed and tested at RIT, to meet predefined usability metrics: (i) effectiveness (accuracy and completeness), (ii) efficiency (resource usage) and (iii) satisfaction (comfort and acceptability of use). The next phase of the project will take place in Kaduna, Nigeria at the the DSDC, where the S-DELI NGO will host the project. They will assist with the user training of parents at the school, assisting them in setting up the tool as an interpretation device, monitoring the deployment of the tools to the home and managing the data collection for further refinement.

At the DSDC, we are currently in discussions with the principal to identify 20 initial families of children ages 3-7 year old, in the early stages of language usage, where their vocabulary is still limited. We will enroll their parents and family members in the first round of the study. Currently, DSDC runs a weekend ASL tutorial session for one hour, where parents and other family members especially siblings of the DHH students are invited and taught basic ASL. The goal is to encourage continued communication beyond the time in school, for the DHH student. This project will leverage such existing processes.

DSDC will conduct quarterly sign language tests (sign speed and accuracy, sign recognition rates, etc) for parents using to the tool, compared with a control matched goup not using the tool. Participants in the study will be given prizes intermittently, to encourage compliance and ensure participants continue to use the tool throughout the duration of the project. Surveys, questionnaires and on-site interviews will be given to measure the quantity and quality of communication taking place at home with the DHH child.

## 3 Technical Proposal : Two-way Neural ASL Translations

We aim to develop robust neural models capable of faithfully translating sign language to gloss[3] as well as in the reverse direction, from gloss to sign, a more challenging AI problem. Sign-to-gloss translation is getting more attention in the AI community , unlike its reverse problem, gloss-to-sign production. To complete the process, we will also train gloss-to-text and text-to-gloss translation models. Generated texts can be converted to speech using open-source speech generators and spoken words can be converted to texts using commercial ASRs (automated speech recognizer).

For word representations, we will use the `fastText`[4] word embedding, a faster and more compact Facebook extension of Word2Vec[5]. Although `fastText` does not consider word ordering, a trait generally considered negatively, in this work, in gloss recognition and production, not paying special attention to word ordering could prove to be beneficial since the grammatical rules of ASL are different from those of spoken English.

For this research, we will conduct the following activities:

1. Create a new multicultural word-level ASL dataset involving subjects from Nigeria, using the sign language version of the Swadesh list [Emmorey and Lane, 2013]. This new dataset will be combined with the existing word-level, 2000-word ASL dataset [Li *et al.*, 2020], comprising of over 21,000 video samples isolated signed words.
2. Extract non-identifiable 3D features from the signing videos, to be released publicly, under creative commons licenses.
3. Optimize our current gloss-to-sign neural model;
4. Apply the dual machine learning paradigm to sign-to-gloss and gloss-to-sign tasks, so that the notion of duality can aid in the learning of each task.
5. Train the auxiliary gloss-to-English and English-to-gloss translation models, to extend the existing sign-gloss system using the continuous, fully-annotated ASL dataset, `How2Sign` [Duarte *et al.*, 2021] comprising of 35,000 video samples of continuous signed phrases annotated with glosses.
6. Test the end-to-end process on ASLing [Ananthanarayana *et al.*, 2021b], our own diverse, real-life ASLing dataset (consisting of over 1,200 video samples of signed phrases), collected in uncontrolled, unconstrained everyday situations at RIT. See Figure 1 for sample frames.

*The novel contributions of this research thrust include the exploring innovative loss functions to improve the gloss-to-sign translation, taking advantage of word embedding*

[3]Gloss is the transcribed form of sign language, which includes various notations to account for the facial and body grammar involved in the signs, but unfortunately, not all signs have a direct meaning in the spoken equivalent.

[4]https://github.com/facebookresearch/fastText

[5]https://code.google.com/p/word2vec/

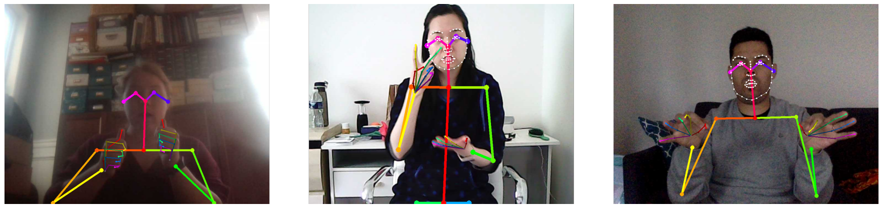


Figure 1: A few sample frames from our real-life ASLing video dataset. Note how some images have poor lighting, the signer's distance to the camera is not fixed, and the presence of noisy, unconstrained backgrounds.

*arithmetic to recognize out-of-vocabulary words and using a dual learning approach to enhance both source-to-target and target-to-source translations.*

### Past Related Works in Sign Production

Neural machine translation (NMT) [Bahdanau *et al.*, 2014], that addresses the translation of one spoken language to another has been the precursor to computational sign language understanding. Rastgoo *et al.* [Rastgoo *et al.*, 2021a] provide an extensive survey of current neural-based sign recognition or sign-to-gloss methods.

But given the significant advances reported for the sign-to-gloss/text recognition task, we focus our research here more on sign production, or gloss-to-sign generation, a more challenging AI problem. One of the earlier sign production models, Text2Sign [Stoll *et al.*, 2020], involved the use of Generative Adversarial Networks (GANs) to generate sign videos. Zelinka et al. [Zelinka and Kanis, 2020] created word-level signs using Openpose [Cao *et al.*, 2021] sequences. Saunders *et al.* [Saunders *et al.*, 2020] generated 3D continuous signs and also created a multi-channel approach using mixture density networks [Saunders *et al.*, 2021]. A GAN based method was used to learn co-articulations in [Saunders *et al.*, 2022].

From our previous works in sign language modeling [Ananthanarayana *et al.*, 2021c; Ananthanarayana *et al.*, 2021a; Wilkins *et al.*, 2020], we developed various architectures for sign language recognition, which we will use in this research. In our most recent work [Ananthanarayana *et al.*, 2023], we developed a continuous text-to-sign model and more recently, we presented a first approach at dual learning for 2-way sign language translation [Chaudhary *et al.*, 2023]. Rastgoo *et al.* also provide a survey of current neural-based sign production or gloss-to-sign methods [Rastgoo *et al.*, 2021b].

### Current State of Computational Sign Language Research

To explore the extent of sign language understanding as an area of technical research, we mined text data from IEEE Xplore digital library[6].

Although other technical document libraries exist, these IEEE Xplore numbers show the trend in the research. Figure 2 shows the progression of the total number of sign language related technical publications through the years, obtained by searching IEEE Xplore for titles containing the phrase *sign language*. The dark blue bars in the chart indicate the number

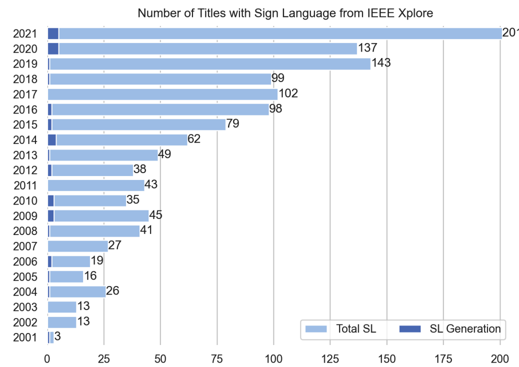



Figure 2: Chart of the numbers of sign language (SL) publications per year from IEEE Xplore

of publications (within those returned for the sign language) that contain the word *production*, *generation* or *animation*, which have to do with sign generation, a main focus of this work. The highest number of papers published was **201**, in 2021, compared with **5,860** - the number of titles with the word *speech* in the same year; the highest number for sign generation was **5**.

Sign language as a technical AI research area is only just budding. This is likely due to recent advances in deep learning technologies that have resulted in significant improvements in spoken language translation [Sutskever *et al.*, 2014; Vaswani *et al.*, 2017], video captioning [Venugopalan *et al.*, 2015; Aafaq *et al.*, 2019], visual storyboarding [Chen *et al.*, 2019] and gesture recognition [Karpathy *et al.*, 2014; Pigou *et al.*, 2018]. There have been many other works in these areas, and these are a few of the pivotal works cited.

**Text-to-sign Generation Model** Although the generation of continuous ASL from its spoken counterpart using neural models, continues to make progress, its rate of success is slow compared to how fast individual ASL sign recognition is progressing [Adaloglou *et al.*, 2022]. We therefore propose a new 2-stage approach to address text-to-sign language translation, where the first phase involves a text-to-gloss translation, and the second stage will involve single word to sign translation.

Figure 3 shows the proposed architecture where the two neural translation models are variants of the traditional encoder-decoder transformers [Vaswani *et al.*, 2017]. We will initially train the model on the 2000 words from *WLASL* [Li *et al.*, 2020] and other word-level datasets.

[6]The IEEE Xplore is a resource for scientific and technical content providing web access to more than five million publications in electrical engineering, computer science, and electronics.

**So why not just have a lookup table for 2000 words?** By training the model to specifically learn the mappings between word embeddings and signs, we can extrapolate signs from unseen words, under the assumption that words that are close in the word embedding space will be close in the sign space.

Motivated by our previous work in continuous sign language generation, we introduce a hybrid loss consisting of the traditional regression loss and a novel triplet loss for recognition.

**$L_2$ Regression loss ($\mathcal{L}_a$)** The objective here is to learn the probability $p(V|W)$ of producing a sequence of pose frames $V = (s_1, \ldots, s_T)$ over $T$ time steps, given a spoken/written language word $W$.

Given the text word embedding $W$ as input, the decoder will output a sequence of pose frames that can be expressed as $\hat{s}_{1:T} = \hat{s}_1, \ldots, \hat{s}_T$. The Mean Squared Error (MSE) loss between the predicted sequence, $\hat{s}_{1:T}$, and the ground truth,$s^T_{1:T}$ is given as: $\mathcal{L}_a = \mathcal{L}_{MSE} = \frac{1}{T}\sum_{i=1}^{t}(s^T_{1:T} - \hat{s}_{1:T})^2$

**Sign similarity metric-based loss ($\mathcal{L}_b$)** For this loss, we are interested in ensuring that the pose frames predicted by the architecture are as similar to the ground-truth signs as possible, and as distant as possible to other signs in the same training batch. To accomplish this, we have:

$$\underbrace{\|f(B) - f(T)\|^2}_{d(B,T)} - \underbrace{\|f(B) - f(S)\|^2}_{d(B,S)} \leq 0 \tag{1}$$

where $B$ is a baseline sign, $T$ is a truth sign required to be as similar to $B$ as possible and $S$ is a false sign (not as similar to the baseline); $d(.)$ is the distance function. To avoid a trivial solution where our function $f(.) = 0$ , or $f(B) = f(T)$, we define a distance function such that
$d(B, T, S) = \max(\ (d(B, T) - d(B, S) + \alpha), 0\ )$
We refer to the loss derived based on this distance as the *sign similarity metric-based loss function*.

The sign similarity based loss over all $M$ samples can thus be given as: $\mathcal{L}_b = \sum_i^M d(B^{(i)}, T^{(i)}, S^{(i)})$

**Total loss:** The overall architecture is therefore trained using a weighted combination of the regression and metric-based losses: $\mathcal{L}_{Word2Pose} = \lambda_a L_a + \lambda_b L_b$. . We will also explore the Connectionist Temporal Classification (CTC) loss, given that the ordering of the predicted and ground-truth signs would be similar.

*The novelty here arises from the fact that the proposed model considers the input word embedding during training, therefore, for inference, if it encounters a word that it is not familiar with, as long as that word is close enough in word embedding space to another word that the model has encountered, it will generate a sign to potentially preserve the meaning of the seen word, given the novel one.*

**Dual learning for 2-way translation** We will use a modified version of our previously developed sign-to-text model [Ananthanarayana *et al.*, 2021a] (the primal task), which at the time was SoTA algorithm on the RWTH PHOENIX-Weather-2014T benchmark dataset [Necati Cihan Camgöz, 2018] and combine this with our new gloss-to-sign model described above (the dual task), in a a dual learning mechanism motivated by [He *et al.*, 2016].

If we define a sign phrase as $\mathbf{x}$ and its textual translation as $\mathbf{y}$, then for a bilingual sign-text sentence pair $(\mathbf{x}, \mathbf{y})$, ideally $p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x})p(\mathbf{x}|\mathbf{y}) = p(\mathbf{y})p(\mathbf{y}|\mathbf{x})$.

If the two models are only trained apart, it becomes challenging to satisfy $p(\mathbf{x})p(\mathbf{x}|\mathbf{y}) = p(\mathbf{y})p(\mathbf{y}|\mathbf{x})$; but joint training of the two models can be performed as:

$$\begin{aligned}\mathcal{L}_{DL} = (&\log \hat{p}(\mathbf{x}) + \log \hat{p}(\mathbf{y}|\mathbf{x}; \theta_{x\to y}) \\ &- \log \hat{p}(\mathbf{y}) - \log \hat{p}(\mathbf{x}|\mathbf{y}; \theta_{y\to x})\end{aligned}$$

where we $\hat{p}(\mathbf{x})$ and $\hat{p}(\mathbf{y})$ can be viewed as empirical statistics of the data. We anticipate that this form of joint training will enhance both the sign-to-text and text-to-sign translations.

**Experimental Plan and Asessment of Outcomes**
We will use the annotated continuous datasets to train a text-to-gloss model, and use the isolated sign datasets to train for gloss sign production. We will test the full text-to-sign model on the non-glossed continuous datasets with captions but no gloss. These will be useful for the overarching text-to-sign translation.

Using back translation[7], we will run the generated signs thorough SoTA pose-to-text models and evaluate the resulting phrases via BLEU[8] and ROGUE[9] metrics. Empirically, a translation system is understandable and has good translations when its **BLEU_1 and ROGUE scores are between 30 and 40**; hence, this is our target metric. We will pay particular attention to corresponding BLEU_4 scores, to ensure it falls between 10 and 20 or higher.

## 4 Project Implementation Plan

In this section, we discuss the various stakeholders of the project, the NGO in Nigeria that will be critical to the launch of the project in Nigeria, the details of how to translate the research to practice, and evaluation of the project implementation.

### 4.1 Project Team - Roles and Responsibilities

**Ifeoma Nwogu** works on AI based human behavior modeling and computational sign language understanding [Ananthanarayana *et al.*, 2021c; Ananthanarayana *et al.*, 2021a; Wilkins *et al.*, 2020]. As a faculty member at both UB and RIT, (UB as an Associate Professor in CSE and Adjunct Professor in Computer Information Science PhD School at RIT), she is very familiar with research processes at both institutions. Her prior work in modeling dynamical systems relating to human behavior will be well suited for this research.

[7]Back translation is the process of re-translating content from the target language back to its source language in literal terms.

[8]BLEU stands for BiLingual Evaluation Understudy, a metric that evaluates how good a translation is by comparing the predicted text to its ground truth equivalent. BLEU 1 - BLEU 4 scores evaluate the performance based on 1-gram (individual words) to 4-gram words (group of consecutive four words).

[9]ROGUE stands for Recall-Oriented Understudy for Gisting Evaluation [1]. It is a set of metrics that compare an automatically produced summary or translation with human-produced references.

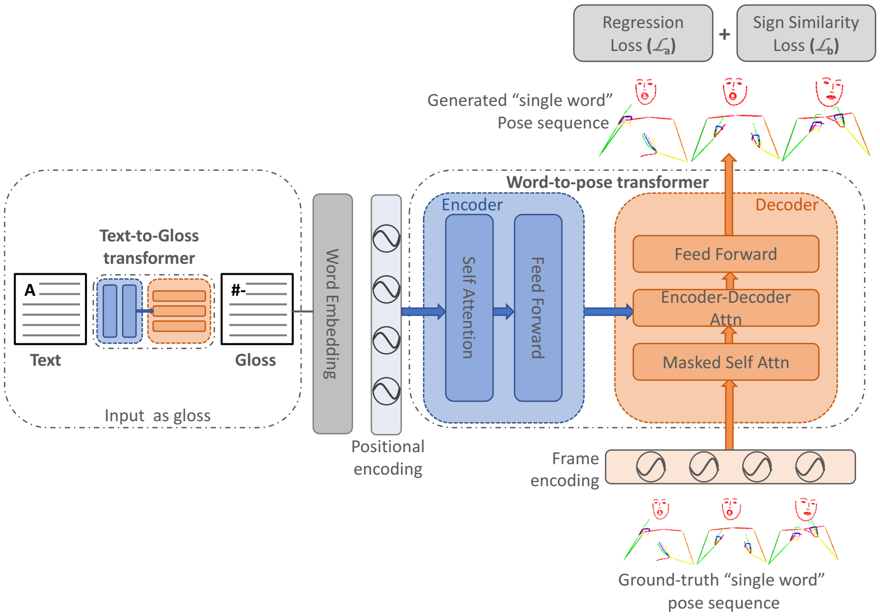



Figure 3: Architecture details of our sequential text-to-sign generation process, showing the novel text-to-gloss preprocessing step on the left and the gloss-to-pose model trained with the hybrid losses.

Nwogu will be responsible for the overall project coordination both in the US and in Nigeria.

**Roshan Peiris** is a faculty member in the Golisano College of Computing and Information Sciences at RIT. His research focuses on human computer interaction, including designing, developing and deploying technologies for disabled individuals especially haptics-devices for visually impaired individuals [Teh *et al.*, 2008]. For this project, Peiris will be responsible for the front-end usability design of the interpretation device. He will also lead the initial efforts in requirements gathering.

**Karthik Dantu** is a faculty member in the CSE at UB, and has worked extensively on perception and coordination in multi-robot systems and mobile systems. This includes multi-modal Visual SLAM [Adhivarahan and Dantu, 2019], permanence reasoning, resource constraints in Visual SLAM and semantic mapping. He was part of the team that helped write the five-year US roadmap in robotics [Christensen *et al.*, 2021] in 2020. For this project, Dantu will work closely with the AI team on the timely realization of the device.

**Ruchi Gamta** is currently a faculty member at NTID is a member of the Deaf community. She has done research in Special Education, Science Education and Secondary Education. Gamta was previously an Instructor in various science subjects at the Model Secondary School for the Deaf in Washington, D.C. Given her background in this area, her role in this project will be that of a subject matter expert.

**Emmanuel Asonye** is a Visiting Research Scholar at University of New Mexico. By training, Asonye is a Field Linguist specializing in the studying and documenting endangered sign languages around the world. He has also done extensive research to better understand the Deaf culture in Nigeria [Asonye *et al.*, 2018]. Asonye is the CEO of S-DELI, the advocacy NGO in Nigeria, slated to host the project. For this project, he will assist in the data collection in Nigeria as well provide the platform to execute the project successfully there.

**Victoria Adesina** is the Principal of the DSDC in Kaduna, Nigeria, the case study for the proposed project. She has been with the school since 1999. Having being trained as an Educator from the University of Port Harcourt Nigeria, and taking a special interest in ASL, Mrs Adesina has worked with DHH children in the elementary and early secondary schools across Nigeria for over 30 years. She is a major stakeholder in this project, after having shared her vision of the DHH students being provided with the opportunity to communicate extensively beyond their time in school during the day.

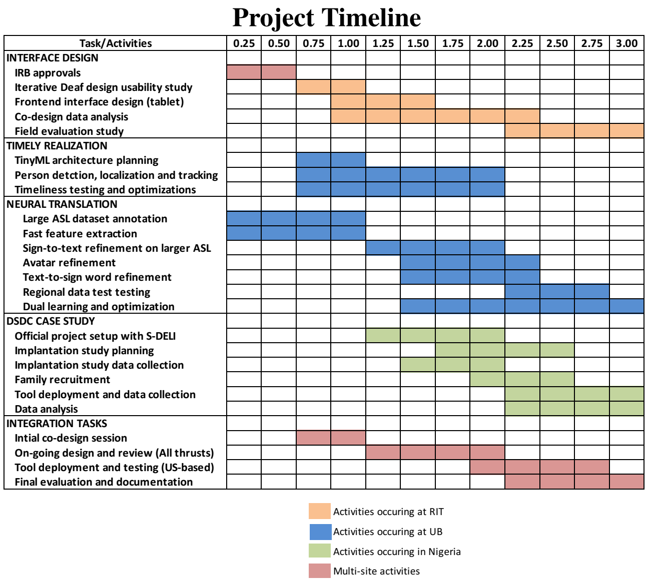
## 4.2 S-DELI: Nongovernmental Organization

Save the Deaf and Endangered Languages Initiative (S-DELI) a community based nongovernmental organization (NGO) committed to the development of Deaf children and adults in Nigeria, founded by co-Author Emmanuel Asonye, a Field Linguistic whose research focus is on documenting endangered indigenous languages, with a special focus on indigenous sign languages in Nigeria. The organization consists of 16 active members and an Advisory board drawn from academic linguists and public school teachers and a computer scientist and other community activists focused on the empowering Deaf children in Nigeria.

Founded in 2018, S-DELI has embarked in six major Deaf student related projects involving state governments and other government parastatals in Nigeria. The most prominent recent project was the Indigenous Nigerian Sign Language Documentation Project (INSLDP) where the a 12-person team from the Kaduna State government, was trained for two days on sign language documentation. Dr Emma Asonye along with the President of the Nigerian National Association of the Deaf (NNAD) were among the trainers at the event. The role of S-DELI was to logistically support the project, providing accommodation (or lodging apartments) for the duration of the project, transportation to and from the collection sites, recording and storage equipment, paraphernalia such as t-shirts, etc. The S-DELI team facilitated any specific regional logistics requirements related to the project. Findings from the project will be used to in Dr. Asonye's research in indigenous sign linguistics. Other projects implemented by S-DELI involve community outreach projects such as the *Beauty Beyond Speech*, a campaign launched in 2017against all forms of sexual and domestic abuse against deaf girls in Africa, while advocating for their safety and education.

We expect to engage with S-DELI in a similar capacity as with the INSLDP project.

### Project Tasks and Timeline




## 4.3 Link from Research to Practice

### Current State of the Technical Application

- **Sign-to-text (S2T):** For this branch of the project, we have currently implemented an extensive sign language translation application, the only known continuous ASL translation project in the literature in [Ananthanarayana *et al.*, 2021b]. For this work, the current BLEU1 score was 22.39 and BLEU4 was 12.25. While these numbers appear low for real-life deployment, the limitation is due to the limited sizes of existing annotated continuous ASL datasets. When a similar algorithm was applied on the annotated German sign language dataset, RWTH PHOENIX-Weather-2014T benchmark dataset [Necati Cihan Camgöz, 2018], which is about 8x larger than our ASLing dataset, the resulting BLEU scores were 5x and 3x, respectively. A larger training dataset will significantly improve the current S2T metrics.
- **Text-to-sign (T2S):** For this significantly more challenging branch, we have implemented a commercially available speech-to-text product for spoken English speech transcription. We are currently able to successfully produce up to 2000 words from the ASL word-level dataset [Li *et al.*, 2020], where these words overlap and are a significantly larger superset of the Swadesh list [10]. Speaking users are restricted to single words in this first iteration, due to the limitation of the technical solution which is currently only able to produce single words. The usability testing will determine whether the application can still be deployed in spite of this limitation.
- **Avatar development for sign production:** Animating avatars for sign language comprehensibility was studied in detail by [Kipp *et al.*, 2011], where they describe methods for the creation and evaluation of a signing avatar. They described techniques useful for enhancing comprehensibility in a short development time and also argue for future research to focus on non-manual aspects and prosody, in order to reach the comprehensibility levels of human signers. We will continue to iterate our signing avatar development, in a similar progression as laid out in the paper. Currently, we have a multi-racial, multi-gender, single word signing avatar, that displays facial expressions, though not yet coordinated with the manual signs. Feedback from the usability study will guide the next iterations of the avatar development.
- **Usability test:** A first usability test will be conducted once IRB approvals are obtained both at UB and RIT, for the series of human studies, including the study in Nigeria. The co-design usability study will include Deaf students, their hearing, non-signing peers, ASL interpreters, Deaf student instructors, and the computing graduate students involved in developing the solution. The goal of the studies will be to determine the aspects of the application that is ready for deployment while identifying others that we consider to be `showstoppers`. A showstopper is a user-requirement identified during testing without which the application can be successfully deployed. Some attributes to be tested will include the comprehensibility of the translation (either way), the appearance of the avatar, its scale and realism while signing, the ease of operating the tablet to be used, the optimal distance of a signer from the tablet camera, etc. Investigators Peiris and Gamta at RIT will lead the co-design efforts in Rochester, NY.
- **First implantation test:** The first experiment to be conducted from Kaduna, Nigeria, which we refer to as the *implantation test* will involve the collection of about 200 signed phrases from the local signers in Nigeria. This set of continuously signed phrases should capture the universal concepts for the regional sign language, and will be compiled by our collaborating Nigerian linguists including Dr. Asonye. After incorporating the feedback from the usability test, we will test the regional data set to observe the resulting BLEU1-4 scores. They should be in a similar range as the US-based ASL translation scores. Similarly, we will obtain a set of signed word videos of 200 words similarly identified by the Nigerian linguists. We will expect to obtain an accuracy of greater than 75% on the 200 regional words. We anticipate that this implantation test will highlight some deficits in the US-trained system which will have to be addressed before a full at-home deployment is conducted.
- **Tool deployment:** The tool will be deployed with the assistance of S-DELI and DSDC in Nigeria. From the DSDC, we will identify 20 families of Deaf children 3-7 year old in the early stages of language usage, where their vocabulary is still limited, and enroll their parents/caretakers in the study. Currently, DSDC conducts weekly sign language tutorials (sign speed and accuracy, sign recognition rates, etc) on Saturdays, for students' parents. We will leverage this established communication protocol to deploy the tool and train its users. Surveys will be given to measure the quantity and quality of communication taking place at home with the Deaf child. Ongoing feedback from the users will be incorporated into the research and development process.

### Budgetary Requirements

Programmer to complete interface and GUI (Hourly at $25/hour for 40 hours) - **$1000**

Student researchers to develop the algorithms (1 grad students at $50k/year for 3 years); Hourly students for 500 hours of work per year at $16/hour; for 3 years - **$174,000**

Research Administrator in Nigeria for $40k/year - **$120,000**

ASL Consultant in US $12k/year - **$36,000**

Computational linguistic research and data gathering $25k/year - **$75,000**

[10] The Swadesh list in classical linguistics, is a compilation of potentially universal concepts useful for learning and describing a language.

Data collection costs - videos, surveys, interviews, room rentals, etc. at $3,000 per year - **$9,000**

Tablets for 20 families at $125 per tablet + participatory incentive costs $5 per family per quarter for 3 years - **$3,700**

Travel cost for 2 investigators to Nigeria for 5 days every year at $2500 per person per year - **$15,000**

**TOTAL = $433,700**

Note: Researchers' salaries and fringe benefits not included

### Potential Problems

1. The most prevalent limitation of the project that we anticipate is from the automated interpreting avatar, which is limited in its performance. Sign language consists of manuals and non-manuals, but the proposed interpretation tool is currently only taking manuals into consideration. Manuals are the linguistic features for sign language, which include the hands - its shape, location, orientation and movement [Neff *et al.*, 2008]. Nonmanuals are the linguistic features which include facial expression, gaze and torso movements and these may stretch over several signs. The absence of nonmanuals for this deployment could present a risk in the adoption of the technology. To mitigate this, we have employed facial expressions in the avatars although they are not yet coordinated with the manual signs.
2. Although ASL is the foundational sign language used in teaching students in schools across Nigeria, there are regional "dialects" of ASL, as well as indigenous sign languages used in the home. Hence, attempting to introduce ASL initially trained on American subjects could be a limitation of the project - the language usage, word orderings and sign vocabularies could potentially be different. To mitigate this, as the system continues to be iteratively updated using data obtained from Nigeria, the effects of training originally only on the US data could be reduced over time.
3. The commercially available speech-to-text and text-to-speech converters are trained and designed to work with voice types primarily from US and Europe. It is not clear how readily these systems will be able to understand the accents and other language features in Nigeria. Although the initial sets of users will be family members who speak English, they do so with regional intonations and slang. Similarly, it is not yet known, how readily the regional participants will understand the spoken version of the signed phrases that were translated to text and then converted to speech again by commercial speech converters. To mitigate this, we will explore speech-text converters that can train on the users' voices in their initial stages.
4. The system is designed to translate spoken English to ASL and vice versa. But many DHH student's families, who are the target audience of the interpretation tool, are not native English speakers. Hausa is the main language spoken in Kaduna, the city where DSDC is located. Deploying the tool in homes where English is not spoken conversationally would be a limitation of the project. This can be mitigated by including an English to Hausa and Hausa to English speech translator, such as found in Google Translator. To initially work around this issue, our first cohort of families will be ones where the parents speak English, even if not as a native language.

## 5 Project Evaluation

Over the course of the three years of the project duration, the research team will continue to improve and deploy the tool to meet the defined usability standards in terms of it effectiveness, efficiency and user satisfaction.

The research staff will be trained to employ a basic general language assessment test, the Kendall P-Levels conversation proficiency test[11] [French, 1999]. This will be used assess how well the family members are improving in their sign language development. The test assigns one of eight grades from level `P-Level 0+` (primary forms of communication include informal gestures, facial expressions, and differentiated cries), through `P-Level 1` (the person calls attention to physical needs and expresses personal reactions. The person uses a similar behaviors as 0+, but can now imitate signs produced by others, though not initiating the signs themselves yet) ... to `P-Level 7` (the person can say what she has in mind to say without circling around it, can provide abstract, detailed information such as the rules of a complex game and can use other words to say the same thing so the other person understands, if needed).

## Ethical Considerations

All human research protocols to be implemented in this work will be required to be approved by Institutional Review Board (IRB) before the research can go forward. All research administrators, both within and outside the US will be required to have appropriate certifications to carry out the research work.

During the field evaluation, although the control group will not receive access to the tool at the onset of the project, they will still be invited to participate in the hour-long ASL weekend tutorials offered on the DSDC premises. Families will not be kept as control groups for longer than one year. Lastly, since we plan to collect data from families of children with disabilities, we will doubly ensure that their identities are protected in any video recordings obtained.

## References

[Aafaq *et al.*, 2019] Nayyer Aafaq, Ajmal Mian, Wei Liu, Syed Zulqarnain Gilani, and Mubarak Shah. Video description: A survey of methods, datasets, and evaluation metrics. *ACM Comput. Surv.*, 52(6), October 2019.

[Adaloglou *et al.*, 2022] Nikolas Adaloglou, Theocharis Chatzis, Ilias Papastratis, Andreas Stergioulas, Georgios Th. Papadopoulos, Vassia Zacharopoulou, George J. Xydopoulos, Klimnis Atzakas, Dimitris Papazachariou, and Petros Daras. A comprehensive study on deep learning-based methods for sign language recognition. *IEEE Transactions on Multimedia*, 24:1750–1762, 2022.

[Adhivarahan and Dantu, 2019] C. Adhivarahan and K. Dantu. Wisdom: Wireless sensing-assisted distributed online mapping. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 8026–8033, 2019.

[Al-khazraji *et al.*, 2018] Sedeeq Al-khazraji, Larwan Berke, Sushant Kafle, Peter Yeung, and Matt Huenerfauth. Modeling the speed and timing of american sign language to generate realistic animations. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '18, page 259–270, New York, NY, USA, 2018. Association for Computing Machinery.

[Ananthanarayana *et al.*, 2021a] Tejaswini Ananthanarayana, Lipisha Chaudhary, and Ifeoma Nwogu. Effects of feature scaling and fusion on sign language translation. In *Interspeech 2021, 22nd Annual Conference of the International Speech Communication Association, Brno, Czechia*, pages 2292–2296, 2021.

[Ananthanarayana *et al.*, 2021b] Tejaswini Ananthanarayana, Nikunj R. Kotecha, Priyanshu Srivastava, Lipisha Chaudhary,

[11]https://drive.google.com/file/d/1ePjjTWr2JSTCKMZ8KB8wQLvfO0HT_HTU/view

Nicholas Wilkins, and Ifeoma Nwogu. Dynamic cross-feature fusion for american sign language translation. In *16th IEEE International Conference on Automatic Face and Gesture Recognition, FG*, pages 1–8. IEEE, 2021.

[Ananthanarayana *et al.*, 2021c] Tejaswini Ananthanarayana, Priyanshu Srivastava, Akash Chintha, others, and Ifeoma Nwogu. Deep learning methods for sign language translation. *ACM Trans. Access. Comput.*, 14(4):22:1–22:30, 2021.

[Ananthanarayana *et al.*, 2023] Tejaswini Ananthanarayana, Lipisha Chaudhary, and Ifeoma Nwogu. Signnet: Single channel sign generation using metric embedded learning. In *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)*, pages 1–7. IEEE, 2023.

[Asonye *et al.*, 2018] Emmanuel Ihechi Asonye, Ezinne Emma-Asonye, and Mary Edward. Deaf in nigeria: A preliminary survey of isolated deaf communities. *SAGE Open*, 8(2), 2018.

[Asonye, 2016] E. Asonye. *Sign language: Africa*. The SAGE deaf studies encyclopedia: SAGE Publications, Inc, Thousand Oaks, CA, 2016.

[Bahdanau *et al.*, 2014] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[Bai *et al.*, 2022] Zhen Bai, Elizabeth Codick, Ashely Tenesaca, Wanyin Hu, Xiurong Yu, Peirong Hao, Chigusa Kurumada, and Wyatte Hall. Signing-on-the-fly: Technology preferences to reduce communication gap between hearing parents and deaf children. In *Interaction Design and Children*, IDC '22, page 26–36, New York, NY, USA, 2022. Association for Computing Machinery.

[Best, 2015] Katelyn E Best. "we still have a dream" the deaf hip hop movement and the struggle against the socio-cultural marginalization of deaf people. *Lied und populäre Kultur/Song and Popular Culture*, 60:61–86, 2015.

[Cao *et al.*, 2021] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43(1):172–186, 2021.

[Chapple, 2019] Reshawna L Chapple. Culturally responsive social work practice with d/deaf clients. *Social Work Education*, 38(5):576–581, 2019.

[Chaudhary *et al.*, 2023] Lipisha Chaudhary, Tejaswini Ananthanarayana, Enjamamul Hoq, and Ifeoma Nwogu. Signnet ii: A transformer-based two-way sign language translation model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[Chen *et al.*, 2019] Shizhe Chen, Bei Liu, Jianlong Fu, Ruihua Song, Qin Jin, Li Pingping, Xiaoyu Qi, Chunting Wang, and Jin Zhou. Neural storyboard artist: Visualizing stories with coherent image sequences. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 2236–2244, 2019.

[Christensen *et al.*, 2021] Henrik Christensen, Nancy Amato, Holly Yanco, Maja Mataric, Howie Choset, Ann Drobnis, Ken Goldberg, Jessy Grizzle, Gregory Hager, John Hollerbach, et al. A roadmap for us robotics–from internet to robotics 2020 edition. *Foundations and Trends® in Robotics*, 8(4):307–424, 2021.

[Duarte *et al.*, 2021] Amanda Duarte, Shruti Palaskar, Lucas Ventura, Deepti Ghadiyaram, Kenneth DeHaan, Florian Metze, Jordi Torres, and Xavier Giro-i Nieto. How2sign: A large-scale multimodal dataset for continuous american sign language. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pages 2735–2744, 2021.

[Eleweke *et al.*, 2015] C. Jonah Eleweke, Isaac O. Agboola, and Simon I. Guteng. Reviewing the pioneering roles of gallaudet university alumni in advancing deaf education and services in developing countries: Insights and challenges from nigeria. *American Annals of the Deaf*, 160(2):75–83, 2015.

[Eleweke, 2002] C. Jonah Eleweke. A review of issues in deaf education under nigeria's 6-3-3-4 education system. *Journal of Deaf Studies and Deaf Education*, 7(1):74–82, 2002.

[Emmorey and Lane, 2013] Karen Emmorey and Harlan L Lane. *The signs of language revisited: An anthology to honor Ursula Bellugi and Edward Klima*. Psychology Press, 2013.

[French, 1999] M French. The kendall conversational proficiency levels. *The Toolkit Appendices for Starting with Assessment*, pages 25–42, 1999.

[Grimshaw *et al.*, 1998] Gina M. Grimshaw, Ana Adelstein, M. Philip Bryden, and G.E. MacKinnon. First-language acquisition in adolescence: Evidence for a critical period for verbal language development. *Brain and Language*, 63(2):237–255, 1998.

[Hakiel, 1995] S. Hakiel. A deliverables oriented approach to the integration of hci into the design and development of software products. In *IEE Colloquium on Integrating HCI in the Lifecycle*, pages 3/1–3/4, 1995.

[He *et al.*, 2016] Di He, Yingce Xia, Tao Qin, Liwei Wang, Nenghai Yu, Tie-Yan Liu, and Wei-Ying Ma. Dual learning for machine translation. In *Advances in neural information processing systems*, pages 820–828, 2016.

[Hoff, 2006] Erika Hoff. How social contexts support and shape language development. *Developmental review*, 26(1):55–88, 2006.

[Jaiyeola and Adeyemo, 2018] Mofadeke T. Jaiyeola and Adebolajo A. Adeyemo. Quality of life of deaf and hard of hearing students in ibadan metropolis, nigeria. *PloS ONE*, 13(1):e0190130, 2018.

[Karpathy *et al.*, 2014] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 1725–1732, 2014.

[Kipp *et al.*, 2011] Michael Kipp, Alexis Heloir, and Quan Nguyen. Sign language avatars: Animation and comprehensibility. In *Intelligent Virtual Agents: 10th International Conference, IVA 2011, Reykjavik, Iceland, September 15-17, 2011. Proceedings 11*, pages 113–126. Springer, 2011.

[Leigh *et al.*, 1998] Irene W Leigh, Alan L Marcus, Patricia K Dobosh, and Thomas E Allen. Deaf/hearing cultural identity paradigms: Modification of the deaf identity development scale. *The Journal of Deaf Studies and Deaf Education*, 3(4):329–338, 1998.

[Li *et al.*, 2020] Dongxu Li, Cristian Rodriguez, Xin Yu, and Hongdong Li. Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 1459–1469, 2020.

[McCarthy, 2018] James K. McCarthy. *A shining beacon : fifty years of the National Technical Institute for the Deaf*. RIT Press, Rochester, New York, 2018. SBN 9781939125491.

[McPherson and Swart, 1997] Bradley McPherson and Susan M Swart. Childhood hearing loss in sub-saharan africa: a review and recommendations. *International journal of pediatric otorhinolaryngology*, 40(1):1–18, 1997.

[Necati Cihan Camgöz, 2018] Oscar Koller Hermann Ney Richard Bowden Necati Cihan Camgöz, Simon Hadfield. Rwth-phoenix-weather 2014 t: Parallel corpus of sign language video, gloss and translation. IEEE Conf. on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018, 05 2018.

[Neff *et al.*, 2008] Michael Neff, Michael Kipp, Irene Albrecht, and Hans-Peter Seidel. Gesture modeling and animation based on a probabilistic re-creation of speaker style. *ACM Transactions On Graphics (TOG)*, 27(1):1–24, 2008.

[Pigou *et al.*, 2018] Lionel Pigou, Aäron Van Den Oord, Sander Dieleman, Mieke Van Herreweghe, and Joni Dambre. Beyond temporal pooling: Recurrence and temporal convolutions for gesture recognition in video. *International Journal of Computer Vision*, 126(2-4):430–439, 2018.

[Rastgoo *et al.*, 2021a] Razieh Rastgoo, Kourosh Kiani, and Sergio Escalera. Sign language recognition: A deep survey. *Expert Systems with Applications*, 164:113794, 2021.

[Rastgoo *et al.*, 2021b] Razieh Rastgoo, Kourosh Kiani, Sergio Escalera, and Mohammad Sabokrou. Sign language production: a review. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3451–3461, 2021.

[Saunders *et al.*, 2020] Ben Saunders, Necati Cihan Camgoz, and Richard Bowden. Progressive transformers for end-to-end sign language production. In *European Conference on Computer Vision*, pages 687–705. Springer, 2020.

[Saunders *et al.*, 2021] Ben Saunders, Necati Cihan Camgoz, and Richard Bowden. Continuous 3d multi-channel sign language production via progressive transformers and mixture density networks. *International journal of computer vision*, 129(7):2113–2135, 2021.

[Saunders *et al.*, 2022] Ben Saunders, Necati Cihan Camgoz, and Richard Bowden. Signing at scale: Learning to co-articulate signs for large-scale photo-realistic sign language production. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5141–5151, June 2022.

[Stoll *et al.*, 2020] Stephanie Stoll, Necati Cihan Camgoz, Simon Hadfield, and Richard Bowden. Text2sign: towards sign language production using neural machine translation and generative adversarial networks. *International Journal of Computer Vision*, 128(4):891–908, 2020.

[Sutskever *et al.*, 2014] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112, 2014.

[Teh *et al.*, 2008] James Keng Soon Teh, Adrian David Cheok, Roshan L Peiris, Yongsoon Choi, Vuong Thuong, and Sha Lai. Huggy pajama: a mobile parent and child hugging communication system. In *Proceedings of the 7th international conference on Interaction design and children*, pages 250–257, 2008.

[Treat, 2016] S. Treat. Deaf education: Gallaudet university: How deaf education and special education is being advanced in nigeria. *Retrieved Nov*
, 2:2016, 2016.

[UN Department of Economic and Social Affairs, 2023] UN Department of Economic and Social Affairs. United nations sustainable development goals (sdgs) - the 17 goals, 2023. https://sdgs.un.org/goals.

[Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 6000–6010. Curran Associates Inc., 2017.

[Venugopalan *et al.*, 2015] Subhashini Venugopalan, Marcus Rohrbach, Jeffrey Donahue, Raymond J. Mooney, Trevor Darrell, and Kate Saenko. Sequence to sequence - video to text. In *ICCV*, pages 4534–4542. IEEE Computer Society, 2015.

[Wilkins *et al.*, 2020] Nicholas Wilkins, Max Cordes Galbraith, and Ifeoma Nwogu. Modeling global body configurations in american sign language. In *Interspeech 2020, 21st Annual Conference of the International Speech Communication Association*, pages 671–675. ISCA, 2020.

[Zelinka and Kanis, 2020] Jan Zelinka and Jakub Kanis. Neural sign language synthesis: Words are our glosses. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, March 2020.