**Towards a system to aid communication with Deaf**



Submitted by:

Ahmad Zia Yosfi & Mehreen Najm

Under the supervision of

**Assist. Prof** **Rafiullah Momand**

Department of Information Systems

Computer Science Faculty

Kabul University

Kabul, Afghanistan

**Abstract**

Towards a system to aid communication with Deaf is an experimental system that aims to aid communication between a deaf person and an ordinary person by translating the ordinary person’s speech to sign language. A speech recognizer recognizes speech from the ordinary person and the system then synthesizes the appropriate sequence of signs in American Sign language (ASL) using a specially developed avatar. By using a phrase lookup approach to language translation, which is appropriate for the highly constrained discourse in society, we were able to build a working system that we could evaluate. We summarize the results of this evaluation (undertaken by deaf users and Ordinary people), and discuss how the findings from the evaluation are being used in the development of an improved system.

*Keywords: Machine Learning, Speech Recognition, Machine Translation, Sign Language, Avatar*

**Contents**

1. **Problem Definition**

1.1 Problem statement………..………………………………………………………………………………….1

1. **Introduction**
   1. Background………..……………………………………………………………………………….…………….2
   2. Literature review………..…………………………………………………………………………….……….4
   3. Objectives………….………..…………………………………………………………………………………….4
   4. Dataset………………………………………………………………………………………………………..…….5
2. **Implementation**
   1. **A: prepare the dataset**…………………………………….……………………………………………….6
   2. B: create a tf.data dataset………………………..……………………………………………………….7
   3. C:writing the encoder and decoder model………………………………….…………………….7
   4. D: Define the optimizer and the loss function……………………………………………7
   5. **Checkpoints (Object-based saving) …………………………………………………..7**
   6. **Training ………………………………………………………………………………….....7**
   7. **Translate ……………………………………………………………………………………8**

**4. Conclusion**

3.1 Summary…………………………………………………………………………………………………………….9

3.2 References………………………………………………………………………………………………………….10

**Problem Definition**

**Problem Statement:**

In our modern information and communication society, daily life would be unimaginable without technology. Information and Communications Technology (ICT) is also very useful for people with special needs.

Deaf are people who can’t talk and hear, hearing people are unfamiliar with Deaf because they don’t know their language. They think that if you cannot hear, you can easily access any necessary information by simply reading it in written form, and if you really need to communicate, you can always write your message down. [6]

But this system is not an efficient way to communicate, Deaf have their own language of communication (sign language), they can only understand this language and also they write and read Faster on sign writing.

According to survey in 2015, there are 121 “Deaf sign languages” in the world, but there is not such an efficient mechanism where Deaf can easily understand people thoughts. [7]

**Introduction**

**Background**

People who are Deaf have little or no hearing ability. The word “Deaf”, (often written with a capital “D”) typically refers to people who use Sign Language as a primary language – either directly with others who sign or indirectly through an interpreter with people who do not sign.

Sign language for deaf people has special features that are quite different from those of spoken language. Sign language is an iconic language compared to spoken language, which is more of an arbitrary one. Another crucial difference between the two languages is that sign language does not have its own writing system. Therefore, in order to write descriptions of signs, line drawings, photographs and illustrations have commonly been used, but these represent only a very small moment in the process of actual signing. [1]

They may hear environmental sounds, and may even understand some speech, but they identify with what is known as the “Deaf Culture” [2]. There has recently been considerable research activity in developing automatic systems which can understand and output speech to provide information services or to perform transactions with customers [3]. We have been developing a system which enable people to communicate with Deaf. It is an interactive translation system to assist in the completion of a conversation between an ordinary person and a deaf. The system translates the ordinary person speech into American Sign Language (ASL) and displays the signs using a specially-developed avatar. A comprehensive approach to the task of enabling humans who cannot sign to communicate using sign-language would clearly require the development of a general purpose speech to sign language converter. This in turn requires the solution of the following problems:

1. Automatic speech to text conversion (speech Recognition).
2. Automatic translation of English text into a suitable representation of sign language.
3. Display of this representation as a sequence of Signs using computer graphics techniques.

For many people who have been profoundly deaf from a young age, signing is their first language so they learn to read and write English as a second language [4]. As a result, many deaf people have below-average reading abilities for English text and prefer to communicate using sign language [5].

We will specify how our system will reduce the gap between Deaf and people, which models can we used to automate this system for communication between them and we will describe how technology can facilitate this mechanism.

**Sign Representation: [1]**

A notation system for sign language is strongly needed to advance the study of its structure. Two pioneers, William C. Stokoe, who proposed a notation system for American Sign Language (ASL), and Lynn Friedman, who analyzed ASL from a phonological point of view, worked towards this end. However, their notation systems are rather impractical for general users because they are too technical, much like phonetic alphabets in spoken languages. Therefore, it is necessary to employ another writing system to describe a sign or a signed sentence for everyday purposes. A more suitable sign writing method for this purpose is “SignWriting” by the American movement analyst Valerie Sutton. Her system was applied to Japanese Sign Language (JSL) and the results indicated that the system was effective enough to “write” Japanese Sign Language.

1. Sutton Movement Writing:

Sutton Movement Writing is the International Movement Writing Alphabet (IMWA) used to record all human and animal gestures. The IMWA records the details of movement-based languages.

1. Sutton’s SymbolBank

Sutton’s SymbolBank Database is the official source for all symbols in Sutton Movement Writing it is built into the SignBank computer program.

We used this database as one of our sources for gathering dataset

Sutton’s writing system of sign language is composed with symbols for head and face cues hand shapes, Movements (direction and manner), qualification and punctuation. Signed words written in SignWriting are created by compounding symbols of a highly pictorial design. This system has the following advantages:

1. it is aimed at enabling us to write every sign or signed sentence of every country;

2. because the writing is pictorial, natural shapes and movements of signs can be

realistically shown;

3.face expressions and body movements can be depicted, too.



**Literature Review:**

This chapter is based on research that we have done to solve the problems such as differences between this system and other similar systems already been built. In this system, we tried our best to overcome challenges we observed in current systems. So according to our researches and findings, there were not any system to convert text/speech or ordinary people into sign language and represent it using an avatar. Briefly, there are systems for Deaf like Tessa but they are limited to a Post Office and they used British Sign language (BSL) instead of American Sign Language (ASL).

**Area 1: TESSA**

TESSA is an experimental system that aims to aid transactions between a deaf person and a clerk in a Post Office. These systems are designed to provide translation of conversational speech between languages with a potentially very large vocabulary. It is an interactive translation system but it operates in a very restricted domain and is designed to assist in the completion of a transaction between a Post Office (PO) clerk and a deaf customer. The system translates the clerk's speech into British Sign Language (BSL) and displays the signs using a specially-developed avatar. [8]

**Area 2: HandTalk**

Founded in 2012, Hand Talk performs digital and automatic translation into the Brazilian Sign Language. HandTalk, which is internationally awarded and a reference in the segment, is run by a friendly virtual interpreter, Hugo, a 3D character that makes communication interactive and easy to understand. <https://www.handtalk.me/sobre>

**Objectives:**

We will suggest an efficient mechanism/system that Deaf can easily communicate with ordinary people to understand their thoughts. As Deaf is a part of society and they need to make a connection with people around them. Our system get the text/speech from smart phone, convert it to sign language using machine translation and represent the speech using 3D character. We will use International Sign Writing as an intermediate

**Dataset:**

The dataset used is a sample of common words and sentences that deaf are using daily.

The font that we used is on American Sign Language write in international sign writing.

We have searched many sources for gathering dataset such as Wikipedia pages, some books like (Goldilocks & the Three Bears in American Sign Language, The Book of Jonah) , Signbank and other sources, but the dataset that we used as a first source it is a subset of a much larger dataset, as described in <http://www.signbank.org/signbank.html> . the dataset in this website was so diffused so we gathered and sorted dataset first. In this website there are more than 10000 words in ASL and has the following feature vectors:

𝠀񀀁񀀉񈗥񈗵񋸦𝠃𝤝𝤨񀀁𝤎𝣤񀀉𝣰𝣮񈗥𝤏𝤇񈗵𝣱𝤑񋸦𝤆𝤜 Come

𝠀񀀁񆇡񈗥񋾡𝠃𝤘𝤰񋾡𝣴𝣵񀀁𝣹𝤒񈗥𝣣𝤇񆇡𝤋𝤘 Disappointed

𝠀񀀁񆕁񇆥񆡁񆡁񍘡𝠃𝤧𝤺񆡁𝤜𝤲񍘡𝣴𝣵񆡁𝤐𝤲񆕁𝤔𝤔񇆥𝤎𝤢񀀁𝣹𝤑 Red

𝠀񀀁񆨡񋲡񍦁𝠃𝤟𝤬񍦁𝣴𝣵񀀁𝣾𝤎񆨡𝤓𝤚񋲡𝤓𝤤 Throw in mouth

𝠀񀀙񈁑񋾡𝠃𝤘𝤵񀀙𝣝𝣲񋾡𝣴𝣵񈁑𝣟𝤗 He

𝠀񀀙񈁑񋾡𝠃𝤘𝤵񀀙𝣝𝣲񋾡𝣴𝣵񈁑𝣟𝤗 She

𝠀񀀡񀀁񆫡𝠃𝤛𝤚񀀁𝤄𝣼񆫡𝤉𝣲񀀡𝣱𝣼 111

𝠀񀀡񂤁񁳑񈙇񇆥𝠃𝤯𝤡񀀡𝣝𝣬񂤁𝣱𝣶񁳑𝤆𝣫񈙇𝤑𝤇񇆥𝤈𝤓 155

𝠀񀀡񂤁񁳑񈙇񇆥𝠃𝤯𝤡񀀡𝣝𝣬񂤁𝣱𝣶񁳑𝤆𝣫񈙇𝤑𝤇񇆥𝤈𝤓 One hundred fifty five

𝠀񀀡񂤁񁳑񈙇񇆥𝠃𝤯𝤡񀀡𝣝𝣬񂤁𝣱𝣶񁳑𝤆𝣫񈙇𝤑𝤇񇆥𝤈𝤓 One-hundred-fifty-five

𝠀񀀡񂤁񃋁񀭡𝠃𝤰𝤖񀀡𝣜𝣷񂤁𝣰𝤁񃋁𝤅𝣹񀭡𝤙𝣷 163

Till now for our first training, we gathered 1500 small sentences.

**Implementation:**

As we motioned the development of a general purpose speech to sign language converter requires the solution of the following problems:

1. Automatic speech to text conversion (speech Recognition).
2. Automatic translation of English text into a suitable representation of sign language.
3. Display of this representation as a sequence of Signs using computer graphics techniques.

By assuming “international sign writing” as the most suitable sign representation, we try to found all sources that we can use to gather “English – sign writing” combination sentences for our dataset.

In first try we gathered 1500 words and small sentences combinations, and we use it to build our first experimental model.

We use a sequence to sequence (seq2seq) Neural Machine Translation model to automate the translation job using machine learning technics and After training the model we were able to input an English word, such as \*"¿translator ", and return the sign:

\*" 𝠀񀀁񀀩񆙡񋎩񋎽񂈁񂈉񆿅񆿕񋸥𝠃𝤨𝥇񆙡𝣪𝣟񀀁𝤅𝣕񀀩𝣿𝣚񋎽𝣥𝣱񋎩𝣦𝣆񂈁𝤚𝤌񂈉𝤁𝤌񆿅𝤛𝤮񆿕𝤀𝤮񋸥𝤍𝥁 "

Which is the representation of word “translator” in American Sign Writing.

# to train our model We used Tensor flow in “google colab” and we followed steps in “Neural Machine Translation with Attention” Notebook available at:

# <https://colab.research.google.com/github/tensorflow/docs/blob/master/site/en/r2/tutorials/text/nmt_with_attention.ipynb>

the steps are as follow:

**A: prepare the dataset:**

We stored the dataset in a “.txt” file in Unicode format, every line in this file is an “English + American sign language” sentence and they are separated with tab “/t”

Example:

𝠀񀀁񀀉񈗥񈗵񋸦𝠃𝤝𝤨񀀁𝤎𝣤񀀉𝣰𝣮񈗥𝤏𝤇񈗵𝣱𝤑񋸦𝤆𝤜 ¿come

1. Add a \*start\* and \*end\* token to each sentence.

2. Clean the sentences by removing special characters.

3. Create a word index and reverse word index (dictionaries mapping from word → id and id → word).

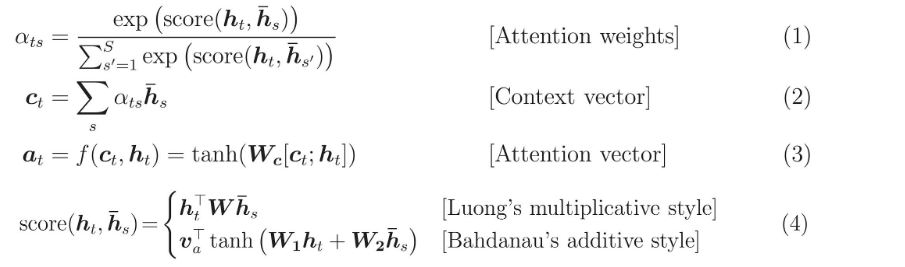
4. Pad each sentence to a maximum length.

### **B: Create a tf.data dataset**

## **C: Write the encoder and decoder model:**

The input is put through an encoder model which gives us the encoder output of shape (batch\_size, max\_length, hidden\_size) and the encoder hidden state of shape (batch\_size, hidden\_size).

Here are the equations that are implemented:



If we consider this notation:

* FC = Fully connected (dense) layer
* EO = Encoder output
* H = hidden state
* X = input to the decoder

The pseudo-code:is as follow:

* score = FC(tanh(FC(EO) + FC(H)))
* attention weights = softmax(score, axis = 1). Softmax by default is applied on the last axis but here we want to apply it on the *1st axis*, since the shape of score is *(batch\_size, max\_length, hidden\_size)*. Max\_length is the length of our input. Since we are trying to assign a weight to each input, softmax should be applied on that axis.
* context vector = sum(attention weights \* EO, axis = 1). Same reason as above for choosing axis as 1.
* embedding output = The input to the decoder X is passed through an embedding layer.
* merged vector = concat(embedding output, context vector)
* This merged vector is then given to the GRU

## **D: Define the optimizer and the loss function:**

## **E: Checkpoints (Object-based saving)**

## **F: Training:**

1. Pass the input through the encoder which return encoder output and the encoder hidden state.
2. The encoder output, encoder hidden state and the decoder input (which is the start token) is passed to the decoder.
3. The decoder returns the predictions and the decoder hidden state.
4. The decoder hidden state is then passed back into the model and the predictions are used to calculate the loss.
5. We Used teacher forcing to decide the next input to the decoder.
6. Teacher forcing is the technique where the target word is passed as the next input to the decoder.
7. The final step is to calculate the gradients and apply it to the optimizer and backpropagate.

## **G: Translate:**

* The evaluate function is similar to the training loop, except we don't use teacher forcing here. The input to the decoder at each time step is its previous predictions along with the hidden state and the encoder output.
* Stop predicting when the model predicts the end token.
* And store the attention weights for every time step.

Note: The encoder output is calculated only once for one input.

## **H: Restore the latest checkpoint and test manually**

We tested the translate function for following words and we get these results:

**Conclusion**

**Summary**

We conclude that the dataset is very small. Our goal in developing this trial system was to establish whether the introduction of a limited speech-to-sign translation system for ordinary people would be beneficial to deaf whose primary means of communication was sign language. Moreover, for testing first version of our system we used American Sign Language, in future it is possible to fit any sign language with our system because we used International sign writing as sign language representation that supports all available languages in the word. In addition, concatenation of signing is more fluent and controlled for avatar than for video signing, as the exact positioning of the avatar can be manipulated. For these reasons, we decided to display the signs using an avatar (3D character).

In next semester we will try to create a large dataset that will cover most of frequently used English word; we will also work on animating the avatar to react properly against every “sign writing” sentence.

**References**

*[1] Kato, Mihoko. "A study of notation and sign writing systems for the deaf." Intercultural Communication Studies 17, no. 4 (2008): 97-114.*

*[2] B. Mazor and B. L. Zeigler. The design of speech-interactive dialogs for transaction automation systems. Speech Communication,*

*17:313–320, November 1995.*

*[3] R. Conrad. The deaf school child. Harper and Row, 1979.*

*[4] Michael Saar and Helena Arthur‐Okor (2013) "Reference services for the deaf and hard of hearing"*

*[5] Maiorana-Basas, Michella, and Claudia M. Pagliaro. "Technology use among adults who are deaf and hard of hearing: A national survey." Journal of deaf studies and deaf education 19.3 (2014): 400-410.*

*[6] Al-Fityani, Kinda, and Carol Padden. "Sign language geography in the Arab world." Sign languages: A Cambridge survey (2010): 433-450.*

*[7] D. Wood, H. Wood, A. Griffiths, and I. Howarth. Teaching and talking with deaf children. John Wiley and Sons, 1986.*

*[8] Cox, S., Lincoln, M., Tryggvason, J., Nakisa, M., Wells, M., Tutt, M. and Abbott, S., 2002, July. Tessa, a system to aid communication with deaf people. In Proceedings of the fifth international ACM conference on Assistive technologies (pp. 205-212). ACM.*