CROSS MODAL TRANSLATION SYSTEM FROM SPEECH IN ENGLISH TO AMERICAN SIGN LANGUAGE

(PROJECT FOR THE COURSE OF COMPUTATIONAL LINGUISTICS, WS 2016-17)

SUBMITTED TO:
Dr. Prof. ALEXANDER KOLLER
DEPT. OF COMPUTATIONAL LINGUISTICS AND PHONETICS,
SAARLAND UNIVERSITY GERMANY

SUBMITTED BY:
HARSHITA JHAVAR (2566267)
MASTERS IN COMPUTER SCIENCE (2016-2018(EXPECTED)),
SAARLAND UNIVERSITY GERMANY

PROBLEM STATEMENT:

Sign language serves as a primary mode of communication for the deaf people. It is a major linguistic challenge which many researchers have taken up and built systems to facilitate the process of understanding the natural language by means of sign language. This document discusses a system that is built mainly to take speech input in form of dialogues in English which are normally spoken while teaching Mathematics to school children. The system generates corresponding American Sign Language Translation of these dialogues as output. The dictionary is limited to 538 sign animations currently which are taken from the Texas ASL mathematics dictionary.

WHY IS THIS PROBLEM IMPORTANT?

The Sign Language is primary way of communication for deaf people. It helps them to communicate by gestures rather than speaking. If a vocal person is unaware of sign language, the communication between hearing impaired people and vocal person is then, mostly achieved through a sign interpreter. It is always not possible to have an interpreter when a person is unaware of sign language. So it becomes highly important to develop a system which can be generates signs for a particular word. With advancements in computers and technology it has become possible to develop an automatic interpreter which can be used to convert Speech to Text which can then further be translated to Sign Language Gestures and can be accessed globally. This software can thus be a good initiative to abridge the gap between the hearing impaired and the other people of our society, leading to more opportunities and exposure for the deaf.

IMPLEMENTATION DETAILS:

As shown in figure 1, the architecture of the system consists of the following components:

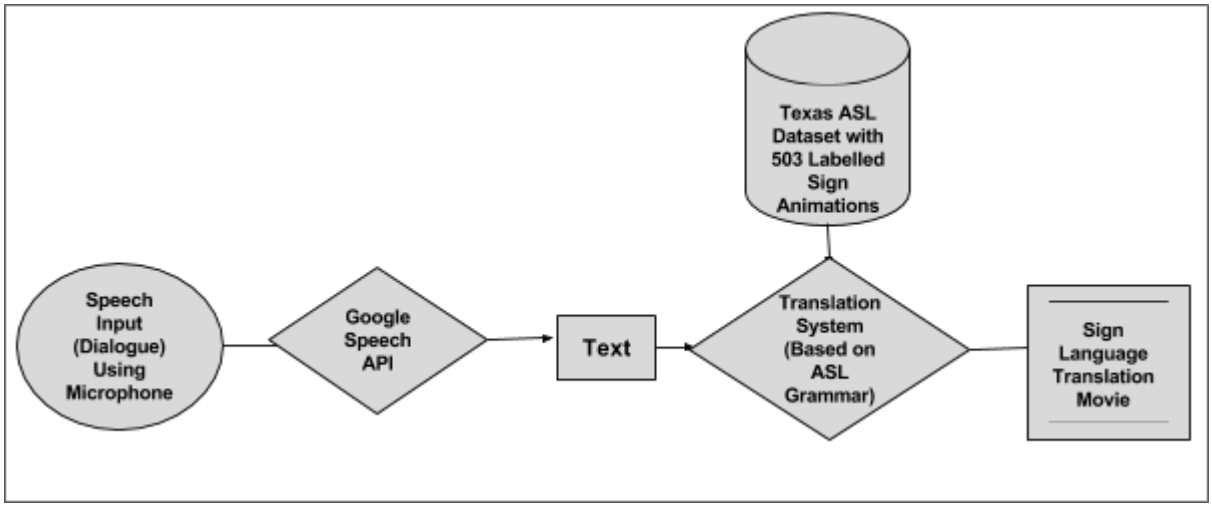


Figure1: A diagram showing the architecture of the current system.

1. **Speech Input using Microphone**

The system becomes more user friendly when it is Speech to Sign Language Translation System. This can be used to teach small hearing impaired kids who cannot read text. This can very well ease the conversation gap, creating more job opportunities and education opportunities for the hearing impaired as the requirement of the knowledge of sign language is no longer a criteria to teach hearing impaired after this software develops further.

The system requires the following to take input in form of dialogue (currently limited to the Texas Mathematical dictionary set) and convert it to text:

- Python 3.3+ (required)
- PyAudio 0.2.9+ (required to use microphone input)
- Microphone
- SpeechRecognizer 3.6.3

Use the following commands to prepare your system for the source code to run successfully on it:

```
git clone http://people.csail.mit.edu/hubert/git/pyaudio.git
cd pyaudio
sudo python setup.py install
sudo apt-get installl libportaudio-dev
sudo apt-get install python-dev
sudo apt-get install libportaudio0 libportaudio2 libportaudiocpp0 portaudio19-dev
sudo pip3 install SpeechRecognition
```

2. **Speech to Text Translation Using Google API**

This API recognizes the input dialogue in English and translates it to Text. The input has to be given by a microphone in a loud clear voice. The dialogue currently can be anything which comprises of the concepts covered in 538 animations of the sign language provided in the dataset as discussed in point 3. Once, we get enough sign animations, the system can be further extended to the level of the daily non-formal interactions in different languages. For example: *“The equation of circle with centre at origin and radius equals to two is x squared plus y squared equals to four.“*

A document titled as ‘sample_dialogues.txt’ in my submission folder consists of 15 such example sentences.

3. **Texas ASL DataSet**

The Texas ASL mathematics dictionary consists of American Sign Language animations of the given 538 words from the mathematical domain. The dialogues currently have to come from this domain only. All the videos are labelled with their English translation. The user is required to use only those words for explaining the concept in simple English sentences. The domain dictionary is available in my submission folder in file 'Dictionary.txt'. One can in future, use English wordnet to improve the scalability of the system.

4. **Translation System based on ASL Grammar**

American Sign Language has its own grammar system, separate from that of English. ASL grammar has its own rules for phonology, morphology, syntax, and pragmatics. Some simple ASL grammar rules are as follows:

-> ASL follows several different "word orders" depending on what is needed. Normal conversations tend to follow **Subject-Verb-Object** order. One might confuse this as to mean *English* word order. While the word order in ASL and English can be similar, ASL does not use BE verbs (am, is, are, was, were) or anything to indicate the state of "being" or the function words, nor does it use articles (a, an, the). Thus, in a sentence like:

"A triangle has three sides."
English: Subject (A triangle), Verb (has), Object (three sides.)
ASL: Subject (triangle), Verb (NULL), Object (three sides)

-> Sign allows you to put the subject before or after the verb when dealing with simple sentences; it doesn't matter which word comes first.
-> In ASL, the direct object can go either before the subject or after the verb.
-> In ASL, we place the indirect object right after the subject and then show the action.
-> Time-Subject-Verb-Object or Time-Subject-Verb word order is used when discussing past and future events. The time-frame or tense is expressed in this order.
These are some rules which are used for general structuring of an ASL sentence. Limiting to the scope of this project, the domain is mathematical terms limited to 538 sign animations as mentioned in the 'Dictionary.txt' file. The input dialogues, sample of which are available in the 'sample_dialogues.txt' file, are assumed to be simple sentences and not, complex and compound sentences.
This system drops the function words whose corresponding translations are not available in the dataset and returns the translation as explained in rules above.
The time_frame is assumed to be present tense currently which is generally the case while teaching elementary mathematics to school kids.
The system checks for the labels (length 1, length 2) in the dataset corresponding to the different words of an input sentence retrieved after dropping the function words.

5. **Sign Language Translation Movie as Output**

The following packages are required to enable the script to merge the relevant videos of sign language animations: PyGame and Moviepy.
Install PyGame using the following command:

python3 -m pip3 install --user pygame

Install MoviePy using the following command:

(sudo) pip3 install moviepy

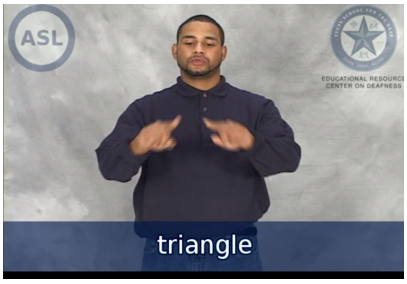
To run and test the system, please make the submission folder as your current directory and run:

python3 source_Speech_To_ASL.py *(For testing Speech to Sign Language Translation)*
(Your system should be connected to a good internet connection for Speech Recognition)
(Please be slow, very loud and clear to give speech input in a silent environment using a good microphone)

Python3 source_Text_To_ASL.py *(For testing only Text to Sign Language Translation)*

EXAMPLES:

1. “A triangle has three sides.”



+


+


= Translation

2. ” The perfect square of two is four.”

+


+


= Translation

3. “The shape of square has four sides.” (Note the difference of the meaning of square here than the previous example.)

+


+


+


= Translation

Other sample dialogues can be seen from the file, “sample_dialogues.txt”

EVALUATION:

For evaluation of this system I showed the results of my system to two scientists: Dr. Alexis Heloir and Dr. Fabrizio Nunnari who work with special focus on Sign Language Synthesis and Interaction at German Research Center for Artificial Intelligence (DFKI, Deutsches Forschungszentrum für Künstliche Intelligenz) in Saarbrücken, Germany.The snapshot of the email on the evaluatory comments on my system after a one hour meeting is below:

Gmail **Harshita Jhavar <harshitajhavar@gmail.com>**

Short review of the demonstration you showed us today.

1 message

Alexis Heloir <alexis.heloir@dfki.de> Wed, Mar 29, 2017 at 3:08 PM
To: Harshita Jhavar <harshitajhavar@gmail.com>
Cc: Fabrizio Nunnari <fabrizio.nunnari@dfki.de>

Dear Harshita,

Here is our review, feel free to paste it in your report.

Best,

Alexis.

Thank you very much for the nice demonstration you gave us today in the seminar room at the CS building (E1.1, room 121). We were pleased to discover a running system consisting of three elements : Speech recognition, feature selection and video playback.

Speech selection is performed using the Google API, it accepts a vocabulary of size 540 lexical items and performed reasonably well. The resulting syntactic tree is then used to retrieve a sequence of motion clips in a video database consisting of 600 clips. The search is a standard label-based search. The average length of a motion clip is five seconds (according to Harshita). clips are played sequentially on the computer screen.

We could interact with the system which was robust enough to stand a series of voiced requests by two different subjects. Naturally we tried to challenge it by providing non trivial requests. The system managed to fail gracefully. To sum up, it was an pleasant and refreshing experience. As a genuine language, Sign Language(s) convey explicit rules at the articulatory, phonetical, phonological and morphological level. Accounting for these rules would greatly improve the current system.

CHALLENGES TO BUILD A CROSS MODAL TRANSLATION SYSTEM BETWEEN ENGLISH TO ASL AND FUTURE DIRECTIONS:

1. Coming up with huge dataset which contains sign animations corresponding to millions of words spoken in English.
2. Coming up with a system which generates sign language, may be trained on a geometric model from the existing sign animations using Machine Learning or Deep Neural Networks.
3. Using Wordnet to improve the scalability of the system to retrieve animations of synonyms of the existing words in the dictionary.
4. Use word similarity to find similar labels for those words whose sign animation is not available.
5. Coming up with Grammar for ASL in two formats:
    a. The Geometrical Grammar based on Spatial Coordinates in a given animation as ASL consists of usage of the nearby spatial geometry or space to express a word. Also, the expressions of eyes, gestures from hands are all included in a geometrical grammar. One can attempt to come up with a Neural Network Model which learns these rules from the existing Sign Animations.
    b. The Syntactic Grammar based on the rules discussed before in the document and concretely, defining the different Non-Terminals like Emotions (happy, sad), Degree of emotions (very happy, sad, grumpy, depressed), Tense to come up with Syntactic rules of ASL grammar as ASL grammar has more inclusion from these than words unlike English.

6. Using parallel corpora of different languages to further make Sign Language Translation Systems in all these languages and thus, helping the hearing impaired community globally because deafness never implies dumbness. They deserve equal opportunity like a normal human being and thus, be an asset to the society, especially in Science and Innovation industry.

BIBLIOGRAPHY:

[1.] **Intelligent Thai text ? Thai sign translation for language learning**
Srisavakon Dangsaart, Kanlaya Naruedomkul, Nick Cercone and Booncharoen Sirinaovakul
Journal: Computers & Education, 2008, Volume 51, Number 3, Page 1125
DOI: 10.1016/j.compedu.2007.11.008

[2.] **HamNoSys generation system for sign language**
Rupinder Kaur and Parteek Kumar
Conference: 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Year: 2014, Page 2727
DOI: 10.1109/ICACCI.2014.6968333

[3.] **Analysis, preparation, and optimization of statistical sign language machine translation**
Daniel Stein, Christoph Schmidt and Hermann Ney
Journal: Machine Translation, 2012, Volume 26, Number 4, Page 325
DOI: 10.1007/s10590-012-9125-1

[4.] **Collecting and evaluating the CUNY ASL corpus for research on American Sign Language animation**
Pengfei Lu and Matt Huenerfauth
Journal: Computer Speech & Language, 2014, Volume 28, Number 3, Page 812
DOI: 10.1016/j.csl.2013.10.004

[5.] **Building machine-readable knowledge representations for Turkish sign language generation**
Cihat Eryi?it, Hatice Köse, Meltem Kelepir and Gül?en Eryi?it
Journal: Knowledge-Based Systems, 2016, Volume 108, Page 179
DOI: 10.1016/j.knosys.2016.04.01