**SPEECH TO SIGN LANGUAGE TRANSLATOR**

**Review video link**

<https://drive.google.com/drive/folders/1hkaiLylLXuBa4CfJhrB6EEcDkFn-fnvY?usp=sharing>

20BCT0216 – Dhruv Sanjay Rao

20BCB0074 – Pranav Bhagwat

20BCE2037 = Shreyskar Shukla

**Abstract**

Communication between deaf and non-deaf people is a difficult task. Only a very small portion of the population is able to communicate with the impaired ones. Unfortunately, there aren’t a lot of systems which are available to the public to help in this communication. Sign language recognition is one of the most growing fields of research today and it is the most natural way of communication for the people with hearing problems. Earlier authors have discussed various machine learning and natural language-e processing methods for speech to sign conversion. Now we propose to use a library, known as mel spectrogram which has been proven to be state-of-art on other benchmark dataset but has not yet been applied for sign language generation. In this paper we present a component to translate English Natural Language to Indian Sign Language

**Keywords**

Sign language recognition, Mel spectrogram, natural language processing, librosa, Symbolic communication, Classification, Tokenization, Hearing impaired, Deaf people

**Introduction**

Sign Language is the main communication method used by the deaf, mute, and hard of hearing individuals.[Fattah 2005] It utilizes hand gestures, movement and orientation of fingers, arms, or body, and facial gestures to convey speaker's ideas. In almost all the world, signs are used to represent the letters of the alphabet with which the oral language of a country is written. This is called manual alphabet or finger spelling. Although known since the 5th century BC, sign language is not standardized internationally and each country generally has its own native sign language. There are perhaps three hundred sign languages in use around the world today. The number is not known with any confidence; new sign languages emerge frequently.[Wikipedia 2022] According to the Center for Strategic and International Studies in Washington DC, out of 350 million people living in the Middle East, over 11 million have a disabling hearing loss but lack of standardized Arabic Sign Language has resulted in dearth of qualified sign language interpreters who can serve as medium of communication between the deaf and the rest of the world[csis.org].

Taking a close look at those values, we can easily conclude that a deaf or hard of hearing person will have immense difficulty while trying to communicate with the general population, being that in most cases, that communication will not be possible, since very few people have the knowledge sign language or that SL users can read or interpret written natural languages . All regions of the world also have same problems due to lack of standardization. This language barrier arises because the deaf usually do not master written language, and only a few hearing people can communicate using sign language [Barros 2016], as there are notable differences in grammar, morphology, syntax, and semantics. Learners that were born deaf, typically, face great difficulties during the acquisition of reading and writing skills [Selva et. Al. 2017].

Nevertheless, speaking society is gradually recognizing the importance of unified sign language and efforts are underway to cp.reate awareness in the society to make deaf and mute useful members of the community at large. A sign language interpreter is very hard to find. The ratio of can go as bad as one in ninety three thousand individuals.[Jamil 2018] there is a desperate need to develop an automatic machine-based translator/interpreter that can convert from speech/text to Sign Language and vice versa.[gayyar 2016]

Even when deaf have a problem with the computer and need to access customer service, deaf people face lots of issues to understand the attendant instructions and end up needing family’s or friend’s help. As observed in research conducted in 2019.[Sannomia 2019] and was pointed out in a systematic mapping of the area[Zeledon 2019], deaf people don’t like to feel different from other people and don’t want to have to put on anything extra to start a normal conversation. Most of them will be able to access a service and execute tasks by their selves when they have, at least, some instructions in Sign Language.

The recognition of signed languages differs widely. In some jurisdictions (countries, states, provinces or regions), a signed language is recognised as an official language; in others, it has a protected status in certain areas (such as education). Symbolic recognition does not guarantee an improvement in the lives of signed-language users as these laws are hardly implemented. Some laws state that it is government, family, and whole society's obligation to provide deaf not only their basic rights, such as access to information, job, and education, but also access to technological and scientific advances. [Wikipedia 2022]

Despite that, there are lots of issues regarding the real inclusion of the deaf. The deaf students that helped themselves to be in the school until the high school are too few and highly spread through the school system.[Rumjanek 2019] They usually do not have other deaf in their classes and cannot discuss the topics they have learned in Sign Language. As a result, they also do not learn vocabulary related to some areas, such as science and technology. That will also impact their interest in some jobs and also will pose a barrier to access many customer services related to information technology.

Earlier, paper authors have designed a sensor-based real-time interpreter for Arabic Sign Language which can convert the hand-signs of the Arabic alphabets into text displayed on a dot matrix screen[Jamil 2018] Recent Artificial Intelligence (AI) developments like deep learning and transfer learning can assist in bridging this communication gap, while improving the classification accuracy and computational power needed at the inference phase, adding the value of automation, pattern recognition, feature extraction, and neural learning, reducing costs of software development and maintenance[bengio 2015] .

At present new technologies for the integration of people with disabilities are developed to achieve a “plural” society “without digital divides”. Over time this integration has improved and has achieved greater interaction in the use of mobile devices and disabled persons.[Flavio et. al. 2015]

Continuing with the efforts in the same direction, in this paper, we are presenting an intelligent system which can transform speech to text using state of the art models and then use database comparison for text to sign languageThe notion is to use an existing state-of-the-art library which is LIbrosa and use it to translate natural language to text and then use dictionary based translation for text to sign.

**Literature Survey**

In Related works authors have used Rule based machine translation[Sannomia 2019] in which methods use unknown symbol removal, stop word removal, text standardization, morphological, syntactic, semantic analysis.[Jamil 2018] Then reorder words and map them to database/dictionary.

In **Neural Machine Translation Approach in Automatic Translations between Portuguese Language and Portuguese Sign Language Glosses[**Alves et. al.**]** proposed system translates that text into glosses via a trained Neural Machine Translator. It uses a Feed-forward backpropagation neural network to perform the translations. All properties are grouped in an object with the original word. The object has the word, word stem, pattern, type, gender, number, prefix, and suffix

The objects are encoded to numeric values and given to the Feed-forward back-propagation neuronal network. The output of the network is then given to a decoder, which decodes the numeric outputs to their corresponding gloss. In this method lemmatization is not performed.

The Rule Based Machine Translation (RBMT) focused on the morphological, syntactic, semantic domains, mainly in characteristics of the form or structure, language rules and the basic meaning of the sentence. In the Statistical Machine Translation (STM) analyses existing translations developed by humans. The general difference between RBMT and SMT is that the first approach is a word-based approach and the second are built on phrase-based systems. In parallel the Neural Machine Translation (NMT) differs from the other two approaches by the ability to learn from each translation task and improve upon each subsequent translation.

In **Sign language to text and vice versa recognition using computer vision in Marathi [**Kagalkar 2015**]** authors have discussed image pre-processing and database retrieval. The input image may be either alphabet image, word image, or image of sentence. Whatever the processing is done it is through predefined and form trained database. Input image will be stored in the database as per in required format. Pre-processing includes the selecting input, either image or text. Then grey scale convergence, edge detection, generation of array of image, compare or matching with database. All these functionality comes under pre-processing. If input given to system is correct or in proper way then pre-processing will be done correctly as well as easily. For edge detection Canny’s edge detection algorithm is used.

In **Fine-tuning a pre-trained Convolutional Neural Network Model to translate American Sign Language in Real-time[**Cayamcela 2019**]** authors discuss a new technique to enhance the accuracy from previous approaches by using a pre-trained CNN architecture, re-using the layers with the trained weights on feature extraction. This approach intends to improve the accuracy of actual alphabet sign language recognizers, rather than simply applying feature extraction and image pre-processing. The notion is to modify an existing state-of-the-art CNN by replacing the last set of fully-connected layers with our new set of fully connected layers with randomly initialized weights.

In **w2v-BERT: Combining Contrastive Learning and Masked Language Modeling for Self-Supervised Speech Pre-Training [**Chung 2021**]** explores Masked language model for self-supervised speech representation learning. w2v-BERT is a framework that combines contrastive learning and MLM, where the former trains the model to discretize input continuous speech signals into a finite set of discriminative speech tokens, and the latter trains the model to learn contextualized speech representations via solving a masked prediction task consuming the discretized tokens. w2v-BERT greatly improves a real-world recognition task.

In **Automatic Translate Real-Time Voice to Sign Language Conversion for Deaf and Dumb People[**Gayathri 2017**]** The proposed framework use previously existing instructive recordings on the web and gives sign subtitling accessible during the run of the video. The framework is versatile and simple to use in a study hall setting which can make a useful expansion to help the information base of understudies.

In **Sign Language Recognition using Microsoft Kinect [**Agarwal 2013**]** approach is divided into two major feature capture modules. One module focuses on building the depth profile of the gesture whereas the other module focuses on building the motion profile of the gesture. Through analysis of existing techniques and approaches, It was found that both depth and motion profiles are essential to provide good accuracy.

In **Sign language to speech conversion [**Vijayalakshmi 2016 **]** the proposed system is a sensor based gesture recognition system which uses flex sensors for sensing the hand movements. The flex sensor is interfaced with the digital ports of Atmega328 microcontroller. The output from the microcontroller is the recognized text which is fed as input to the speech synthesizer. Arduino microcontroller processes the data for each particular gesture made. The system is trained for different voltage values for each letter.

In **Automated Speech to Sign language Conversion [**Bharti 2019**]** The approach starts with the acquisition of the input speech which is then con- verted into text using Google API .Information technology is the solution for such problems. In quest to seek a most natural form of interaction, people have promoted the development of recognition systems, e.g. text and gesture recognition systems. For each word/character in the processed text, we then perform matching operation with the visual sign word library(video database of sign language) for successful retrieval of the matched videos.

In **Speech to Indian Sign Language Translator[**Kulkarni 2021**]** proposed model will successfully convert the given input audio into an animation. Many improvements along this route can be made as and when the ISL Dictionary grows. The words in the ISL are small, so many improvements can be made by adding new words to their dictionary to increase their breadth

In **Signed Languages in Natural Language Processing [**Yin 2021**]** the underlying architecture for the systems are based on: Direct Translation, Statistical Machine Translation, Transfer-based Architecture. The system focuses on Indian Sign Language

**Proposed Architecture Diagram**

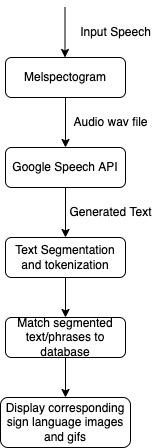
****

Fig. 1: Architecture diagram showing flow

of data from being speech to sign

**Explanation**

**Melspectogram**

The mel scale is a scale of pitches that human hearing generally perceives to be equidistant from each other. As frequency increases, the interval, in hertz, between mel scale values (or simply mels) increases. The name mel derives from melody and indicates that the scale is based on the comparison between pitches.

The mel spectrogram remaps the values in hertz to the mel scale.

The linear audio spectrogram is ideally suited for applications where all frequencies have equal importance, while mel spectrograms are better suited for applications that need to model human hearing perception. Mel spectrogram data is also suited for use in audio classification applications.

**Google Speech API**

Used googles speech API to analyse the audio wav file and detect and extract the text from it to be further used in our program

**Text Segmentation and Tokenisation**

Tokenisation is the process of breaking up the sequence of characters in a text by locating the word boundaries, the points where one word ends and another begins. For computational linguistics purposes, the words thus identified are frequently referred to as tokens. In written languages where no word boundaries are explicitly marked in the writing system, tokenisation is also known as word segmentation, and this term is frequently used synonymously with tokenisation. Text segmentation is the process of determining the longer processing units consisting of one or more words

**Match segmented text to database**

Text is compared with stored database GIFs. If any phrase is matched then it’s GIF is displayed. Else, from alphabet dataset each letter corresponding to text is displayed

**Datasets**

<https://github.com/satyam9090/Automatic-Indian-Sign-Language-Translator/tree/master/ISL_Gifs>

<https://github.com/satyam9090/Automatic-Indian-Sign-Language-Translator/tree/master/letters>

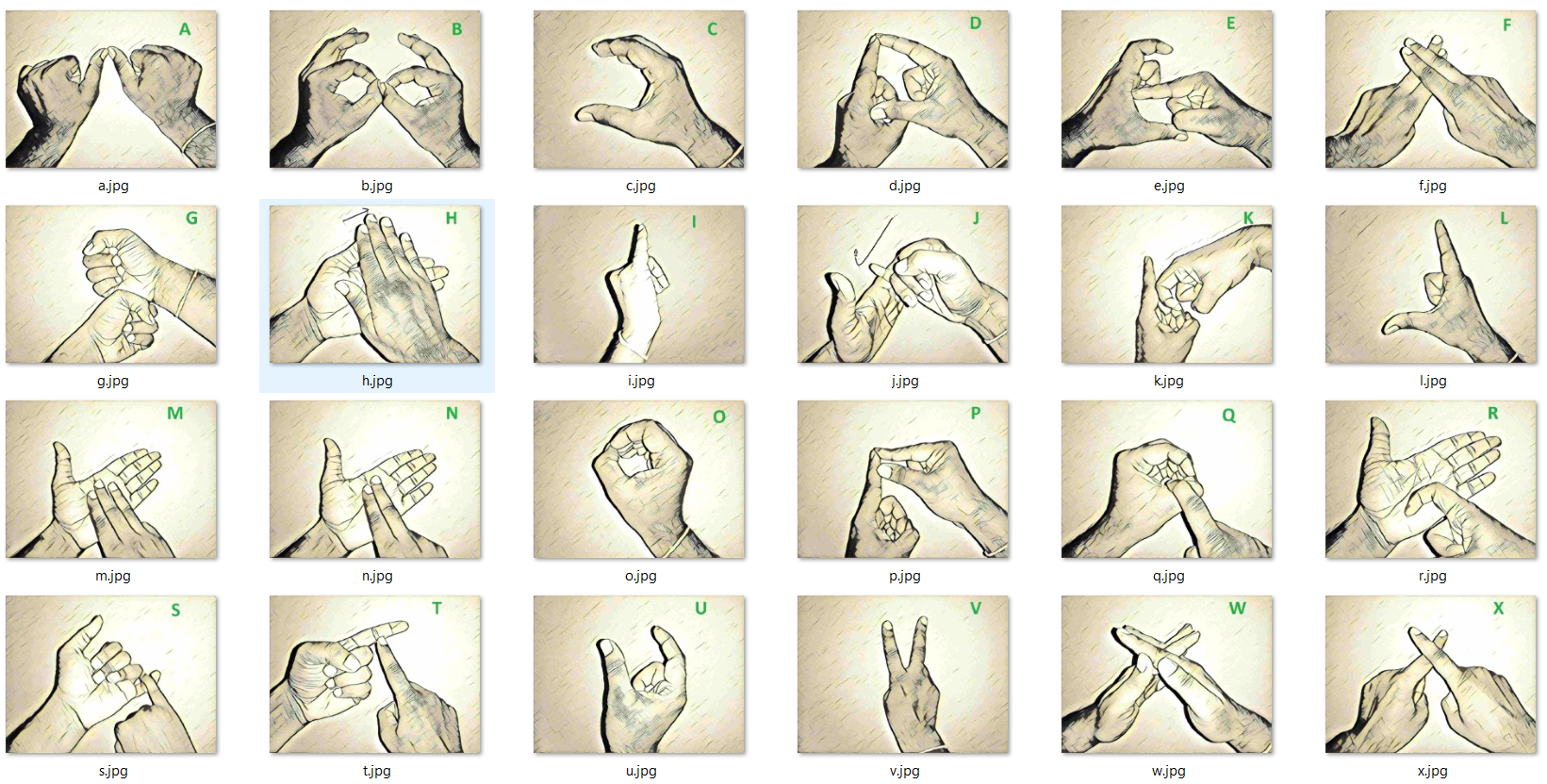


Fig. 2: Screenshot of alphabet dataset being used

We have used datasets which have GIFs of common phrases and other dataset is all 26 alphabets of English language

**Results**

**Input text: hello**

**Output**

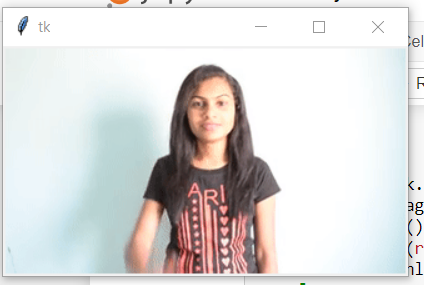


Fig 3. Still from output GIF displaying *Hello*

**Input : GO**

**Output**

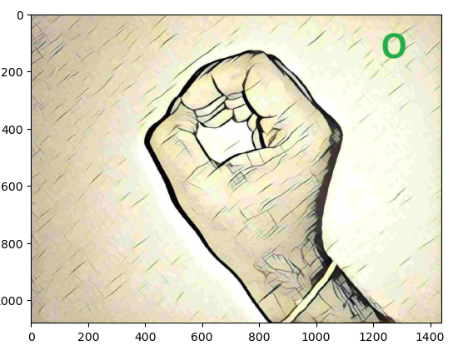
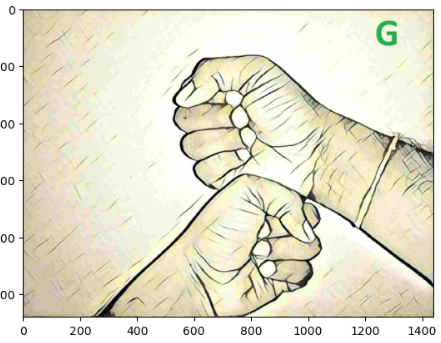


Fig 4: Output images showing *GO*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Serial No. | Test case description | Expected result | Actual Result | Testcase passes(P) or failed(F) |
| 1. | Input speech recognized or not? | Input speech is recognized | Input speech recognized | P |
| 2. | Input speech recognition was accurate or not? | Input speech is recognised accurately. | Input speech is recognised accurately. | P |
| 3. | The text is accurate or not? | Text is accurate | Text is accurate | P |
| 4. | Is the text or sentence meaningful? | The text is meaningful. | The text is meaningful | P |
| 5. | Is the text represented in sign language properly? | The text is represented properly in the sign language. | The text is represented properly in the sign language. | P |
| 6. | Is the listening button working properly? | The button is working properly. | The button is working properly. | P |

**Comparison with other papers**

**BLEU Score**

It rates translation based on its closeness with original sentence. So it gives us accuracy of model.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Paper No.** | **1**[Gayyar, 2016] | **2**[Jamil 2018] | **3**[Peral 2019] | **4[**Shinde 2015**]** | **5[**Gayathri 2017**]** | **6[**Our paper] |
| **Input text** | **BLEU Score** | | | | | |
| How are you | 1.000 | 1.000 | 0.867 | 0.930 | 0.940 | 0.910 |
| I am fine | 1.000 | 1.000 | 0.940 | 0.980 | 0.905 | 1.000 |
| Go | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| Hello | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| I, the author, am not going back | 0.900 | 1.000 | 1.000 | 0.866 | 0.740 | 0.867 |

**Conclusion**

Sign language translator is extremely useful in various areas. We have used Melspectogram to improve the audio quality and google speech API to convert speech to text. This text is then then tokenized and segmented to detect phrases. The tokenized text is compared with the database containing alphabets and gifs on the text in sign language.

**References**

1. *M. A. Abdel-Fattah, "Arabic Sign Language: A Perspective",* Journal of Deaf Studies and Deaf Education*, vol. 10, no. 2, pp. 212-221, March 2005*
2. *Wikipedia contributors. "List of sign languages." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 27 Jul. 2022. Web. 22 Sep. 2022.*
3. Reading the Signs: Diverse Arabic Sign Languages*, [online] Available: https://www.csis.org/analysis/reading-signs-diverse-arabic-sign-languages.*
4. *M. El-Gayyar, A. Ibrahim and M.E. Wahed, "Translation from Arabic speech to Arabic Sign Language based on cloud computing",* Egyptian Informatics Journal*, vol. 17, pp. 295-303, 2016.*
5. *Jamil T, "Design of a Real-time Interpreter for Arabic Sign Language",* Proceedings of the IEEE SoutheastCon 2018*, 19–22 April 2018*
6. *J. D. Barros Junior, "Tradução Automática de Línguas de Sinais: do Sinal para a Escrita",* Ph.D. dissertation*, 2016.*
7. *E. P. da Silva and P. D. P. Costa, "QLIBRAS: A novel database for grammatical facial expressions in Brazilian Sign Language",* Proceeding of the X Meeting of Students and Teachers of DCA / FEEC / UNICAMP (EADCA)*, 2017.*
8. *D. Sannomia, T. B. Ferreira and M. C. Ferreira, "Evaluation of the use of accessibility tools in the information technology industry: a case study",* Proceedings of the 2019 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pp. 261-263, 2019.*
9. *L. Naranjo-Zeledón, J. Peral, A. Ferrández and M. Chacón-Rivas, "A Systematic Mapping of Translation-Enabling Technologies for Sign Languages. Electronics",* Electronics*, vol. 8, no. 9, pp. 1047, 2019.*
10. *Wikipedia contributors. "Legal recognition of sign languages." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 27 Jul. 2022. Web. 22 Sep. 2022.*
11. *V. M Rumjanek and W. S. da Silva, "Ciência para todosƒ",* RBPG Brasília*, vol. 15, no. 34, 2019*
12. *Y. A. LeCun, Y. Bengio and G. E. Hinton, "Deep learning",* Nature*, vol. 521, no. 7553, pp. 436-444, 2015, [online] Available: https://www.nature.com/articles/nature14539.pdf.*
13. *P. L. Flavio, A. Darwin, S. Jenny, M. Derlin and I. Alexander, "Design and implementation of an oximetry monitoring device",* 2015 CHILEAN Conference on Electrical Electronics Engineering Information and Communication Technologies (CHILECON)*, pp. 185-189, 2015*
14. *Shinde, A., & Kagalkar, R. (2015). Sign language to text and vice versa recognition using computer vision in Marathi. International Journal of Computer Applications, 975, 8887.*
15. *Gayathri, A., & Sasi Kumar, A. (2017). Sign language recognition for deaf and dumb people using android environment. International Journal of Current Engineering and Scientific Research (IJCESR), 4.*
16. *Agarwal, A., & Thakur, M. K. (2013, August). Sign language recognition using Microsoft Kinect. In 2013 Sixth International Conference on Contemporary Computing (IC3) (pp. 181-185). IEEE.*
17. *Vijayalakshmi, P., & Aarthi, M. (2016, April). Sign language to speech conversion. In 2016 International Conference on Recent Trends in Information Technology (ICRTIT) (pp. 1-6). IEEE.*
18. *Bharti, R., Yadav, S., & Gupta, S. (2019, June). Automated Speech to Sign language Conversion using Google API and NLP. In Proceedings of the International Conference on Advances in Electronics, Electrical & Computational Intelligence (ICAEEC)*
19. *Kulkarni, A., Kariyal, A. V., Dhanush, V., & Singh, P. N. (2021, September). Speech to Indian Sign Language Translator. In 3rd International Conference on Integrated Intelligent Computing Communication & Security (ICIIC 2021) (pp. 278-285). Atlantis Press.*
20. *Yin, K., Moryossef, A., Hochgesang, J., Goldberg, Y., & Alikhani, M. (2021). Including signed languages in natural language processing. arXiv preprint arXiv:2105.05222.*
21. *Cayamcela, M. E. M., & Lim, W. (2019, February). Fine-tuning a pre-trained convolutional neural network model to translate American sign language in real-time. In 2019 International Conference on Computing, Networking and Communications (ICNC) (pp. 100-104). IEEE.*
22. *Alves, V., Ribeiro, J., Faria, P., & Romero, L. (2022, June). Neural Machine Translation Approach in Automatic Translations between Portuguese Language and Portuguese Sign Language Glosses. In 2022 17th Iberian Conference on Information Systems and Technologies (CISTI) (pp. 1-7). IEEE.*
23. *Kajiyama, T., Endo, R., Kaneko, H., Sano, M., & Shishikui, Y. (2022, June). Sign language image dataset with a hand pose type attribute. In 2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB) (pp. 1-4). IEEE.*
24. *Chung, Y. A., Zhang, Y., Han, W., Chiu, C. C., Qin, J., Pang, R., & Wu, Y. (2021, December). W2v-bert: Combining contrastive learning and masked language modeling for self-supervised speech pre-training. In 2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU) (pp. 244-250). IEEE.*