International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 02 | Feb 2019 www.irjet.net p-ISSN: 2395-0072

Tamil Speech to Indian Sign Language using CMUSphinx language Models

Priyanka B R[1], Arvind Mukund[2]

[1,2]*Student, Dept. of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India*

Abstract - *The aim of our project is to help hearing disabled people be independent in their day-to-day conversations and associations. This project is mainly focused on people who know Tamil and are hard of hearing. This project will use Speech recognition algorithms to convert spoken Tamil to a base language stored by the computer in an intermediate format. It will be implemented with the help of using grammar algorithms transformation (Natural Language Processing). This project would be able to either fully or at least sparsely translate the input phrase spoken by the user and output in an avatar depicting the input in Indian Sign Language.*

Key Words: **Language Models, Word Error Rate, Speech Recognition, Indian Sign Language**

1. INTRODUCTION

It is an everyday hassle that people with a hearing impairment face when it comes to understanding the specifics of what people are saying to them. It could be something mundane such as describing a shirt to having specifics about their food. This is intensified when encountering someone who does not know sign language.

Indian Sign Language is not very commonly taught in India. According to the consensus taken in early 2018, the population of people with a hearing impairment is around six million. However, the number of schools that teach sign language is lesser than five hundred. And the schools are scarcely located that families have had to give up their livelihood to move to big cities in order to access these resources.

Aiming to bridge this gap, a system that converts everyday Tamil speech to Sign Language (explained in detail later) is to be employed using concepts under Machine Learning and Natural Language Processing.

2. PRESENT SYSTEM

Although there is no current system that specifically converts Tamil speech to Sign language, there are similar projects in other languages.

In 2008, a system to convert Spanish speech to sign language was proposed to assist people with a hearing impairment. This system was mainly focused on phrases that are spoken by an official while renewing or applying for an identity card.

There are also mobile applications that can convert speech to sign language. Hand Talk, which has been downloaded for over 1 million times as of 2016, enables the users to record their speech which an avatar then translates into sign language. This application is extensively used in Brazil as it converts Portuguese into Libras which is the Brazilian Sign Language. Another application called iCommunicator helps two-way conversion (i.e.) it can convert speech to text or sign language, and it converts sign language to text which is then read aloud by the application. This application is widely used in many corporations around the US and Canada.

Another project that is being developed is the conversion of the Korean closed captions in weather news channels to sign language that is depicted by avatars.

3. PROPOSED SYSTEM

The problem with the existing system is that it does not exist for Tamil language. Since Tamil is phonologically rich, the present systems do not offer the flexibility to help recognize and interpret it.

The system that is being proposed processes Tamil speech and converts it to Indian Sign Language. Therefore, any user can speak into the system which then processes the contextual meaning and displays the corresponding sign language for it.

The converted sign language will be displayed either in the form of videos or programmed avatars depicted the translated input speech.

4. LITERATURE SURVEY

We examined some of the methodologies that were implemented by the present systems and surveyed the usability of the concepts and algorithms that we wish to implement in our system for maximum optimization.

Paper [1] addresses the fact that English and Tamil have acoustically similar phonemes. The difference phonemes for both English and Tamil are recorded. The English phonemes are then mapped to the closest sounding phoneme in Tamil. This would improve our efficiency rate by using an English based Speech to Text model and adapting it for languages by just incorporating the lexicon for different sounding phonemes. This paper would help us identify the phonemes and help optimize the proposed system.

Paper [2] describes a system that they wish to implement in North Korea where they make use of the closed captioning service provided in weather channels and convert it to Korean Sign Language. This paper uses Parts-Of-Speech (POS) tagger from Pusan National University to build their dictionary. To prevent issues such as unavailability of a word in the Korean Sign Language (KSL), a synonym dictionary is built so that the missing word can be translated using the synonym. The dictionary is built using Yoon et al's KorLex which acts as the Korean WordNet. [2] illustrates the use of different Natural Language Processing (NLP) features to help optimize the system. It also highlights the use of the synonym dictionary to increase the range of speech that can be translated. This paper also uses motion capture to create avatars that demonstrate the translated KSL. It does not incorporate any speech recognition since it only considers the closed captions offered in the channel.

Paper [3] illustrates how the lexicon and the orthographic rules of Tamil Language are written in the form of regular expressions. It uses only finite state operations and how this approach can be implemented to develop a morphological analyzer/generator. It identifies the morphological richness of Tamil compared to other universal languages such as English. It is based on an implementation called the Two-Level Morphology which was successful for other languages. This paper helps us understand the scope of the Tamil Language and the different morphological elements that are incorporated. It makes use of Finite State Transducers and Regular Expressions to create a model. This paper uses a XEROX Toolkit to create their model. This paper tries to cover only the verbs and nouns in the language.

Paper [4] describes a system that converts Kannada speech to text using CMUSphinx. This paper also elaborates on the difficulty in creating a language model for Asian languages when compared to English due to the unavailability of multiple language models. [4] Uses publicly available data of a speech that consists of 1000 sentences giving general information about Karnataka. It was able to identify 2112 unique words in the corpora. Feature extraction has a lot of parameters such as the age of the speaker, the upper and lower frequencies. A dictionary is created consisting of the phonemes. [4] Implements two sets of data: the single user speech context dependent, and multi user speech context independent. For context dependent, a Word Error Rate (WER) of 4.6% was recorded with a sentence error rate of 22.5% compared to the context independent 57% sentence error and WER of 19.8%.

Paper [5] proposes a model that extracts features from a preprocessed image and produces a classifier that improves the accuracy of classification in Tamil Sign Language images. Although this paper describes a model that is a reverse of the project in our proposal, this paper gives us more information about the Indian Sign Language and the nuances that exist within it.

Paper [6] defines different methods of speech recognition over a large vocabulary. This paper analyses a variety of techniques like word based, class based, or subword based using neural network language modelling. To benefit acoustic modelling, spontaneous speech corpora were combined with less spontaneous language. Data for Finnish included 550 speakers who all read 30 sentences and 30 words in varying noise conditions. Estonian acoustic models used 164 training data hours comprising of broadcast conversations, news, lectures, and spontaneous conversation. Around 1.3 million words were recorded. The paper concludes by stating that class based neural network language model (NNLM) is the most accurate for Finnish and Estonian speech recognition when compared to word-based models with a WER of 27.1% in Finnish and 21.9% WER in Estonian.

Paper [7] describes a statistical system that translates Spanish speech to sign language. It presents results of translation from both using an automatic speech recognizer (ASR) and written texts. Using both Finite State Transducers (FST) and Phrase Based Systems, [7] considers two different approaches in the translation to sign language. In Phrase Based Systems, GIZA++ which is an automatic statistical translation system, is used. This tool will help train these models for any pair of languages. Bayes theorem is used to represent the probability that a string LO is generated in a string LD. In FST method, a set of strings is generated by GIZA++ using the training corpus and appropriate alignment methods. This can be inferred as a rational grammar which then becomes the topology for FST. The training pairs of sentences are then built by extending the corpus from the pairs of subphrases with GIZA++ wherein the FST is obtained. The WER that is recorded by using the subphrase mechanism is averaged out to be 31.75% and the WER obtained by using the FST is 28.21%

After analyzing the present systems in other languages and the methodology, we come to the conclusion that CMUSphinx is the least erroneous language modeling method to identify Tamil speech and the phonemes. The parameters such as the frequencies, and the age of the speaker are taken into consideration to aid us in implementing the model.

5. IMPLEMENTATION

Implementation of this project depends on the CMUSphinx project or more specifically the pocketsphinx Application Program Interface (API) which contains a lightweight speech recognition library with a python wrapper. The dictionary and phonemes are created manually by hand and once written we can use the g2p-seq2seq to convert our grapheme (word as they are written) to the phonemes. The dictionary contains of a dict file which we use to transcribe words to their phonetic sounds. A language model to calculate the n-gram for each word contextually or alone. We currently support up to 3-gram. The language model dict file and filler words are used generate the trained model by

extracting features from the input training sounds. Once trained using Baum Welch (BM) filters, we can now use the trained model to decode a given speech input. Although similar sounding words prove to be an issue now due to the limited resources, we can obtain for the Tamil Language.

6. CONCLUSION

As seen in the above paper, and the surveys undertaken, we can ascertain the value of this model and the different aspects of communication it can simplify. After reviewing the past systems, using CMUSphinx for creating the language model is proven to be the most efficient methodology. Therefore, our proposed system aims to use the CMUSphinx language model to implement the speech recognition phase of our project. Communication being a basic feature of humans, we seek to make it equal for all to enjoy.

7. ACKNOWLEDGEMENT

We would like to express our gratitude to Professor Dr. T. Sujithra for her advice and assistance in reviewing this research work.

REFERENCES

[1] Exploiting acoustic similarities between Tamil and Indian English in the development of an HMM-based bilingual synthesiser, Sherlin Solomi VijayaRajSolomon ; Vijayalakshmi Parthasarathy ; Nagarajan Thangavelu, 2017

[2] Avatar-Based Sign Language Interpretation for Weather Forecast and Other TV Programs By Juhyun Oh, Seonggyu Jeon, Byungsun Kim, Minho Kim, Sangwook Kang, Hyukchul Kwon, SMPTE Motion Imaging Journal, 2017

[3] Morphological Analyzer and Generator for Tamil Language, S. Lushanthan, A. R. Weerasinghe, D. L. Herath, 2014 International Conference on Advances in ICT for Emerging Regions, 2014

[4] Kannada Speech to Text Conversion Using CMU Sphinx Shivakumar, K. Aravind, K. G., Anoop, T. V. Gupta, D. (2016). Kannada speech to text conversion using CMU Sphinx. 2016 International Conference on Inventive Computation Technologies (ICICT)

[5] Optimized Classification Based on Particle Swarm Optimization Algorithm for Tamil Sign Language Digital Images, Dr M Krishnaveni, Dr P Subashini, T T Dhivyaprabha, A Sathiya Priya, 2016

[6] Unlimited vocabulary speech recognition with morph language models applied to Finnish, Teemu Hirsima¨ki, Mathias Creutz, Vesa Siivola, Mikko Kurimo, Sami Virpioja, Janne Pylkko¨nen, Computer Speech and Language, 2006

[7] Speech into Sign Language Statistical Translation System for Deaf People, B. Gallo, R. San-Segundo, J.M. Lucas, R. Barra, L.F. D'Haro, F. Fernández, IEEE LATIN AMERICA TRANSACTIONS, VOL. 7, NO. 3, JULY 2009

[8] http://www.cmusphinx.github.io: CMU Sphinx API

[9] https://www.npr.org/sections/goatsandsoda/2018/01/14/575921716/a-mom-fights-to-get-an-education-for-her-deaf-daughters