Bisayan Dialect Short-time Fourier Transform Audio Recognition System using Convolutional and Recurrent Neural Network

Patrick D. Cerna, Rhodessa J. Cascaro, Khian Orland S. Juan, Bon Jovi C. Montes, Aldrei O. Caballero

College of Computer and Information Science, MAPUA Malayan College Mindanao, Davao City, Philippines

Abstract—Speech is a form of oral communication that reinforces thoughts and ideas that have general purpose and meaning. In the Philippines, Filipinos can speak at least three languages. English, Filipino, native language. The Philippine government says the Philippines has more than 150 regional native languages, one of which he says is Cebuano. This research aims to implement automatic speech recognition (ASR) specifically for the Bisayan dialect, and researchers use machine learning techniques to create and operate the system. ASR has served its purpose in recent years not only in the official language of the Philippines, but also in various foreign languages. The required datasets were collected throughout the study to train and build the models selected for the speech recognition engine. Audio files are recorded in waveform file format and contain Visayan phrases and sentences. Audio was captured through hours of recorded audio and process using Tensorflow short time Fourier transform (STFT) algorithm to ensure the accurate representation. In order to analyze the audio data, the recordings were specially converted to digital format, specifically .wav and making it sure all records are uncorrupted with only one channel, and finally have a sample rate of 22050kHz. A data mining process was carried out by integrating CNN layers, dense layers, and RNNs to predict the transcription of speech input using multiple layers that determine the output of the speech data. The researchers used the JiWER Python library, which was used in parallel when evaluating WER. This is because the trained scripted data set contains at least 500 time recordings totaling 61.78 minutes. Overall, the WER output is at best 99.53% and the percentage of records used is acceptable.

Keywords—Bisayan dialect; speech recognition; dense layer; CNN; RNN

I. INTRODUCTION

Automatic speech recognition (ASR) systems use algorithms to convert spoken words into text. Companies like Google Cloud, Microsoft Azure, IBM Watson, and YouTube use ASR systems for education, transcription, and disability support. All of these can provide high quality transcripts. A study showed that YouTube was able to transcribe with a word error rate of 28% compared to manual transcription with a word error rate of 17.4% [1]. However, for this study, we only used high quality FLAC files with little background noise. Poor recording quality and high background noise will result in a high word error rate. Since video data is resistant to audio noise, it can be used for speech recognition. Automatic Speech Recognition (ASR), also known as Automatic Speech Recognition, is the process of converting speech signals into text. The text can be in the form of words, phrases, syllables, or any subword unit [2]. ASR is a subset of Natural Language Processing (NLP) in the knowledge domain of artificial intelligence (AI) and aims to make communication between humans and computers as natural as possible. Its ultimate goal is to make human-computer communication indistinguishable from human-human conversation. A deep neural network (DNN) is a neural network with two or more levels of complexity that uses mathematical modeling to process data, and was used in speech recognition in 2010. DNNs have had great success with automatic speech recognition. However, due to differences in speaking styles and attitudes, models that can account for small changes or perturbations in the feature space lead to overfitting and poor generalization, which is desirable [3].

Accent can be defined as a style of pronunciation in a language. In a Filipino setup, a speaker's accent can be heavily influenced by other speakers in their approximate geographic location. This allows people to speak the same language with different accents, resulting in languages (such as Filipino) that are used with multiple accents. Extensive analysis of this provides information on speaker status, age, gender, dialect, and ethnicity [4]. Recently, the value of recognizing a speaker's accent has begun to be noticed in the field of computers. Its influence has been recognized as a basis for the development of various large-scale speech applications such as automatic speech recognition [5]. However, automatic detection of accents is a challenging research topic because languages can have multiple pronunciation styles. Automatic accent detection (also known as accent identification) is based on the consistency of acoustic patterns that can be identified in speaking styles to identify pronunciations in the same accent cluster.

Previous studies on ASR used hidden Markovs, but some focused only on vowel recognition in Tagalog [6]. Another study by Fajardo et al. [7] Automatic Filipino speech recognition is done using a convolutional neural network (CNN) model with SqueezeNet architecture for Filipino. Meanwhile, E developed continuous speech recognition for Bicol and Kapampangan using the CMU Sphinx Toolkit. The speech corpus was collected by the researcher and consists of seven hours of recordings from 150 native speakers of Kapampangan and Bicolano [8]. A hardware-based speech recognition system [9] is built with circuits incorporating neural networks, including passive and active filters to drive microphones, input/output ports, EEPROM memory, and other

components. Advantages of this type of system include speed, accuracy, and lower cost than software-based speech recognition products.

Cebuano is a native language of Cebu City, of which the Visayan dialect is part. This has been extended not only to the Visayas but also to many places in Mindanao due to its growth and usefulness. There was no formal training in the Visayan dialect during the author's primary and secondary education. However, it is learned through the life stages and experiences of the community and relatives. When the K-12 Enhanced Basic Education was signed into law in the Philippines in 2013, native language subjects were included as part of the language classes for kindergarten and primary school. This gave children insight, depth of knowledge, and understanding of their mother tongue, as well as exposure to English and Filipino from an early age. Just as the Visayan dialect has evolved over the years, like other native languages that have been taught in primary school for almost a decade, these languages may one day gain the upper hand for some reason.

However, not all languages are being supported by this technology due to a very large amount of languages existing in our world, and only a few people are working on it. Thus, the researchers were able to implement Automatic Speech Recognition (ASR) in the Bisaya dialect. By doing so, the researchers would be able to provide a new path for the Bisaya dialect in the implementation of technologies alongside the expanding community in our country. In addition, the researchers would also be able to promote the use of their language in technologies, thus making it popular among the people to reduce the risk of being lost in the future. However, not all languages are being supported by this technology due to a very large amount of languages existing in our world, and only a few people are working on it. Thus, the researchers are planning to incorporate Automatic Speech Recognition (ASR) in the Bisaya dialect. By doing so, the researchers would be able to provide a new path for the Bisaya dialect in the implementation of technologies alongside the expanding community in our country. In addition, the researchers would also be able to promote the use of their language in technologies, thus making it popular among the people to reduce the risk of being lost in the future.

II. RELATED WORKS

Raval and Gajjar [10] conducted a study in an effort towards filling the gap between differently-abled people like deaf and dumb and the other people. The obtained results after extracting background were used for forming data that contained 24 alphabets of the English language. The Convolutional Neural Network proposed here is tested on both a custom-made dataset and also with real-time hand gestures performed by people of different skin tones. The accuracy obtained by the proposed algorithm is 83%. To develop a system that can read and interpret a sign like Amrutha and Prabu [11], one must train it using a large dataset and the best algorithm. As a basic SLR system, an isolated recognition model is developed. The model is based on vision-based isolated hand gesture detection and recognition. Assessment of ML-based SLR model was conducted with the help of four candidates under a controlled environment. The model made

use of a convex hull for feature extraction and KNN for classification that yielded 65% accuracy.

On the other hand, Adithya and Rajesh [12] presents an efficient convolutional neural network (CNN) based model for automatically recognizing fingerspellings in sign languages. The model has been tested on a novel Indian sign language (ISL) fingerspelling dataset as well as a publicly available hand posture dataset, and has obtained promising results. Similary, Qin et al. [13] construct a lightweight sign language translation network. We construct the dataset called CSL_BS (Chinese Sign Language-Bank and Station) and two-way VTN to train isolated sign language and compares it with I3D (Inflated three Dimension). Then I3D and VTN are respectively used as feature extraction modules to extract the features of continuous sign language sequences, which are used as the input of the continuous sign language translation decoding network (seq2seq). Based on CSL-BS, two-way VTN achieves 87.9% accuracy while two-way I3D is 84.2%. Finally, Suardi et al. [14] created a trial of combining CNN models using the Ensemble method has been successfully carried out with the results being able to increase the accuracy value to 99.4%. and proved that using Ensemble can increase the higher accuracy value.

In the Philipines, Bautista and Yoon-Joong [15] describe the development of speech recognition using the Hidden Markov Model Toolkit (HTK) in Filipino only. Modifications were made to some datasets to remove unwanted background noise that is not needed for speech recognition. Therefore, removing these noises may improve the model's performance. A variety of experiments with the models are specified in the author's study, and the accuracies can be compared to conclude which model is the most effective to use in the study. Finally, Laguna and Guevara [16] used a language identification (LID) system in their approach. LID can recognize languages of unknown languages, Tagalog (TGL), Cebuano (CEB), Hiligaynon (HIL), Kapampangan (KAP), Bicolano (BCL), Warai (WAR), Tausugu (TSG) based on their research. Among these languages were pairwise and hierarchical LIDs, yielding average accuracies of 48.07% for seven languages, 72.64% for pairwise and 53.99% for hierarchical. The researchers say the LID system works best for a small number of target languages that are closely related to Filipino.

III. MATERIALS AND METHODS

A. Research Design

The study intends to implement and create automatic speech recognition for Bisayan dialect. To be more specific, the study will produce an output of the transcribed text from the input audio. But, to do so, the researchers had to collect the necessary datasets that will be used throughout the study to train and build the selected models for the speech recognition engine. Researchers adapted the Knowledge Discovery in Databases (KDD) framework to manipulate data in their research. This data mining (DM) framework was adopted for the research, which consists of six distinct phases as shown in Fig. 1, based on the research goal. This will help researchers identify and collect valuable datasets for developing their own neural network models to achieve her Bisayan speech

recognition. Therefore, researchers also use various techniques to help create and develop useful datasets for training models.
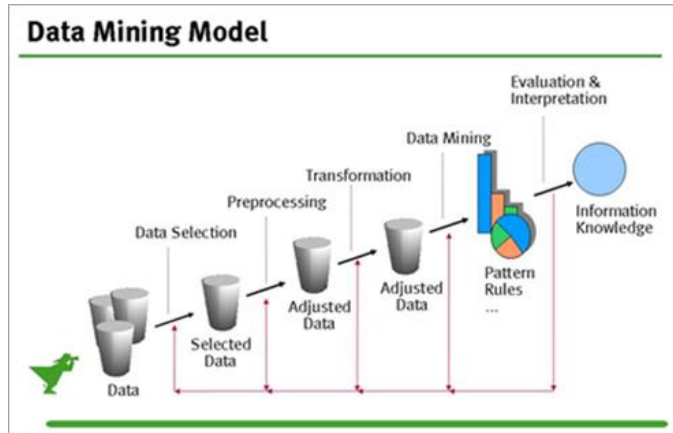

Fig. 1. KDD framework [17].

B. Data Sources and Selection

This study selects research participants who are fluent in the Visayan dialect of the Davao area and collects the required datasets through voice recordings. This helps the authors collect real, authentic data to use in their experiments and analysis throughout the study. These datasets consist of audio files containing unedited recordings of participants' natural voices. This is the raw audio output from the recording device where data collection starts with voice and text data. Audio files are recorded in waveform file format and contain Visayan phrases and sentences. When capturing audio data, hours of recorded audio must be captured to ensure the process is more accurate. During text data collection, everything is controlled and thoroughly researched by researchers to ensure quality and accurate data for research. A phrase or sentence script is provided by the researcher for text data collection (Table I). All these collections are handled according to requirements model and evaluation. So before the actual audio data processing he needs a powerful GPU such as the Nvidia RTX series to process large datasets efficiently and the latest generation of his GTX series performs well I can do it.

Researchers collected speech recognition data from public and private premises with the consent of the participants in Davao City and Mati City, as shown in Fig. 2. Respondents were provided with a script containing random Bisaya phrases of hers. The list of phrases in the Visayan dialect contains hundreds of sentences structured as scripted phrases. Some of these phrases are informal conversations by researchers affiliated with privacy and other illegal matters, but they do not infringe. The recorded phrases are long and short as they provide random topics and fiction for conversation. As shown in Table I, some examples of the following Visayan dialect phrases are read by participating speakers.

C. Methods of Data Processing

The refine input recording audio from raspberry pi powered device which contains waveform file format and this will contain Bisayan phrases or sentences. The captured audio were processed for training and testing through Tensorflow STFT algorithm will use to analyze, synthesize, transform and

describe audio signals and JiWER plugins fo Similarity measures. PRAAT application tool for speech/voice feature extraction will also be utilized as the need arises.

Fig. 2. Audio recording of bisayan speakers.

TABLE I. LIST OF BISAYAN SENTENCES

English	Visayan / Cebuano
How are you?	Kamusta Ka?(Ka-mu-sta ka)
Do you speak english?	Kabolo ka mo sultiogenglish? (ka-ba-lo ka mo salty ogenglish?)
A little	Gamay ra (ga-m(eye) ra)
I speak a little Visayan	Makasulti ko gamay (ma-ka-salty ko ga-may)
I don't understand	Wala ko sabot (wa-la ko sa-bot)
Please speak slowly	Hinaya lang ogstorya (Hi-n(eye)-yah lang og story-ya)
What is you name?	Unsay imongngalan? (Un-s(eye)ee-mongnga-lan)
Thank you	Salamat (Sa-la-mat)
Where can I buy… (item)?	Asa ko makapalit ani (item) e.g shoes (ah-sa ko ma-ka-pa-lit ah-ni)
I would like to buy souvenirs	Ganahan ko mopalit pasalubong (Ga-na-han ko mo-pa-lit pa-sa-loo-bong)
Where is…(place)?	Asa ang lugar (place) e.g CR (term used for bathroom) (ah-sa ang loo-gar)
What places can one visit here?	Unsay ma suroyansainyonglugar? (Oon-s(eye) ma soo-roy-yansaee-mong loo-gar)
How much?	Tag pila? (tag-pi-la)
Good Morning	MayongBuntag (m(eye)-yong bun-tug)
Good Afternoon	MayongHapon (m(eye)-yong hah-pon)
Good Evening	MayongGabii (m(eye)-yong gah-bee-e)
See You Later	Kita ta unya (ki-ta ta oon-ya)
Yes	O-o (oh oh)
No	Dili (Di-li)
What?	Unsa? (oon-sa?)
When?	Kan-usa? (kan-oosa?)
Please	Palihug (Pa-lee-hoog)

The CNN Layer, Dense Layer, and RNN will be used is to predict the transcription of the audio input using several layers that would determine the output of the audio. It is expected that the output of this research will then evaluate the accuracy of the speech recognition by using Word Error Rate (WER), accuracy, precision and recall. Fig. 3 shows the separate steps of a process in sequential order, this is known as Flowchart. The process will start by processing the dataset and getting their audio and text data, which will be used for the training of the model. The audio data shall be used as the input data for the model, while the textual data shall be used as the target data of the model's prediction. While in training, the model shall compare its prediction to the target data, and shall evaluate its errors to modify its weight parameters. During testing, it shall compare its prediction to the target data so that it could compute the Word Error Rate of its output. Thus, the value of the Word Error Rate will serve as its evaluation in transcribing Bisaya Dialect.

To analyze the audio data, the recording was converted into a digital format specifically using audacity in particular into a .wav format. According to Frosch [18], digital audio is presented in many formats, and one of them can also compress audio which is termed as the lossy audio compression format. Lossy sound files compression attempts to reduce the amount of data. In simpler meaning, it allows for even more file size reductions by eliminating certain audio information and simplifying the data. At the same time, the frequency has audio features that are also audio signals that can be used to develop statistical or Machine Learning models. Audio files for this study will contain unedited recordings in a lossy format which will significantly benefit for data compression of the study. All audio recorded was set by 22,050 Hz of Project Rate and Mono Recording Channel to lessen file size and let the model train faster. If some of the recordings do not follow the rule or forgot to set, a code in the Programming section will automatically do it for convenience as shown in Fig. 4.
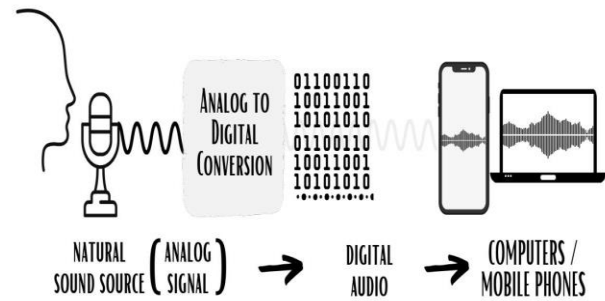

Fig. 4. Analog to digital conversion.

IV. DATA EXPERIMENTATION AND ANALYSIS

A. Audio Conversion to Datasets

After gathering the necessary data for the experimentation, the researchers had to optimize and clean the collected data. Thus, the researchers utilized Python packages to easily optimize and clean their audio dataset so that it would be used for the training of the model. The process was done to ensure that all of their datasets are not corrupted, must be in a .wav format, must not be over 10 seconds, should only have 1 channel, and finally, it should have a sample rate of 22050 kHz. If this would not be followed, it would affect the efficiency of the training process, which would further burden the researchers to do their experimentations. Based on Table II, the researchers had gathered a total amount of 68.76 minutes of audio data that consists of 563 recordings. Each recording contains one phrase out of the 500 phrases that were prepared by the researchers to the participants. The list of created phrases or sentences can be found in Bisaya Dialect Phrases.

TABLE II. DATASETS GATHERED AFTER THE CONVERSION

Group Dataset	No. of Times (mins)	No. of Files
Collected Datasets	68.76	563
Train Datasets	61.78	506
Test Datasets	6.97	57
Augmented Datasets	123.57	1012
Train + Augmented	185.35	1518
Overall	192.33	1575

After optimizing and cleaning the audio datasets, they would be separated into two folders which will be called "train_wav" and "test_wav". These folders shall be the basis for the train datasets and test datasets of the experimentation. After processing the required audio data, the researchers had to create a CSV file for the train and test datasets that would be used to fit the model as shown in Fig. 5. There would be two CSV files that would be generated which are called "train.csv" and "test.csv". These CSVs shall contain the file directory of the audios while also having the text transcription of it. In order for the neural network model (CNN and RNN) to train, the researchers had to extract the audio data and its transcription so that it would be given to the model. The necessary data that needs to be processed is the input and target data for the model. By getting the input data, the researchers use a Python package that is capable of reading the signal from the audio and

Fig. 3. Flowchart of dialect recognition system.

transforming it into input data for the model. For the target data, the characters in the transcription of the audio shall be converted into numbers which will be the target data that the model should have after computing the input data as shown in Fig. 5.
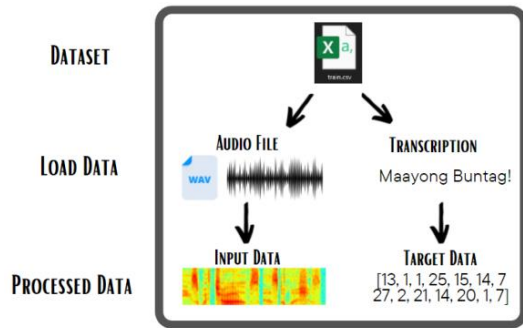

Fig. 5. Audio conversion to target data.

B. Prediction using CNN, RNN and Dense Layer

The CNN Layer, Dense Layer, and RNN were used to predict the transcription of the audio input using several layers that would determine the output of the audio. Based on the figure below, the structure of the neural network model used in this study contains several layers as shown in Fig. 6. The layers within the models consist of several layers. Furthermore, the researchers will use some other underlying layers that are used to optimize the training performance of the model, such as dropouts and normalizations. With this kind of network model structure, the researchers believe that it will be able to perform its purpose and provide results from the given input audio. The researcher will then evaluate the accuracy of the speech recognition by using Word Error Rate (WER), accuracy, precision and recall.
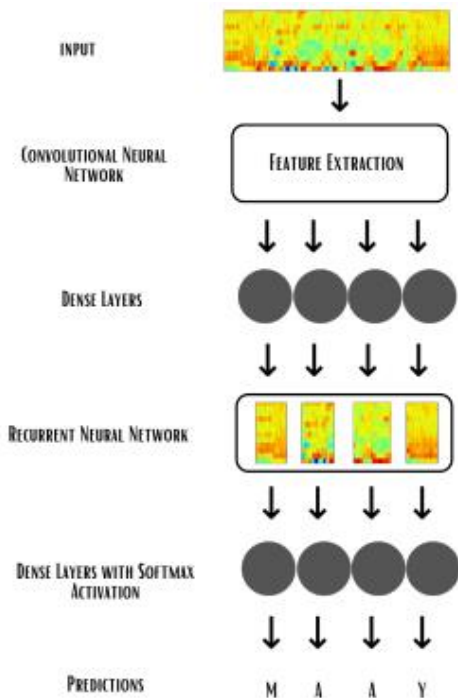

Fig. 6. Prediction of recognition using CNN, RNN and dense layer.

C. Result of Data Analysis

Based on the results of experimentations as reflected in Table III, the researchers believed that adding augmented data would provide better model predictions. Furthermore, having additional RNN layers would also provide positive training results but would make the model predictions perform slower than having a few layers. Thus, depending on the target development and dataset would determine how many layers would be necessary to have better results in the training and deployment. However, since the researchers had only a few datasets to train the model, they had to provide alternative solutions based.

TABLE III. RESULT OF AUDIO RECOGNITION USING CNN AND RNN

Row Number	Audio Dataset	RNN Layers	Optimal Epochs	Word Error Rate
1	11.95 minutes (conversation)	2 layers	100 Epochs	99.53 %
2	11.97 minutes (scripted)	2 layers	92 Epochs	96.37 %
3	11.95 minutes + 11.97 + Noise	2 layers	64 Epochs	98.48 %
4	39.13	5 layers	81 Epochs	76.79 %
5	61.78 Minutes + 123.57 minutes (Augmented)	5 layers	30 Epochs	66.82 %

Furthermore, having additional RNN layers would also provide positive training results but would make the model predictions perform slower than having a few layers. Thus, depending on the target development and dataset would determine how many layers would be necessary to have better results in the training and deployment. However, since the researchers had only a few datasets to train the model, they had to provide alternative based on their experimentations shown in the table above. In the Word Error Rate (WER), this is the standard evaluation of how accurate an Automatic Speech Recognition (ASR) system is. It calculates the number of found "errors" in an ASR transcription text. This has been used by different researchers and big companies worldwide to also measure and identify the accuracy of their machine learning. In this study, the researchers used JiWER Python Library and is used alongside in evaluating WER. Since the trained scripted datasets contains at least 500 and a total of 61.78 minutes' time recording. The WER output shows 99.53% (as reflected in Table III) at best which results to acceptable percentage for the number of datasets used.

V. CONCLUSION AND RECOMMENDATIONS

Overall, the study was able to create and train a neural network model that would be used for Speech Recognition System in Bisaya Dialect. The result of the data experimentation reveals the best results of analysis through 99.53% Word Error Rate for their trained model. The researcher therefore recommends that it would be best to seek ways to address the biggest limitation of the study which is to acquire a large amount of dataset for the training of the model. By doing so, the trained model would be able to have a better prediction and performance, since it has been exposed to a large amount of different data for its training. Aside from that, others should also be thoughtful when it comes to their

hardware specifications, as the training of the model would require a heavy amount of processing power which would put a lot of stress onto your hardware. Thus, it would be recommended to have better hardware when training a neural network model with a very large amount of datasets, so that it would lessen the amount of time to train the model.

REFERENCES

[1] J. L. Bautista and Y. Kim, "[PDF] An Automatic Speech Recognition for the Filipino Language using the HTK System | Semantic Scholar," *[PDF] An Automatic Speech Recognition for the Filipino Language using the HTK System | Semantic Scholar*, Jan. 01, 2019. https://www.semanticscholar.org/paper/An-Automatic-Speech-Recognition-for-the-Filipino-Bautista-Kim/7d47daadfa9738461db5073afaa0f4798baed86e.

[2] D. Nagajyothi and P. Siddaiah, "Speech Recognition Using Convolutional Neural Networks," International Journal of Engineering & Technology, vol. 7, no. 4.6, p. 133, Sep. 2018, doi: 10.14419/ijet.v7i4.6.20449.

[3] D. Guiming, W. Xia, W. Guangyan, Z. Yan, and L. Dan, "Speech recognition based on convolutional neural networks," 2016 IEEE International Conference on Signal and Image Processing (ICSIP), Aug. 2016, Published, doi: 10.1109/siprocess.2016.7888355.

[4] M. Tjalve, Accent Features and Idiodictionaries. Department of Phonetics and Linguistics 2007.PhD Dissertaion. University College London.

[5] A. Mukhamadiyev, I. Khujayarov, O. Djuraev, and J. Cho, "Automatic Speech Recognition Method Based on Deep Learning Approaches for Uzbek Language," *Sensors*, vol. 22, no. 10, p. 3683, May 2022, doi: 10.3390/s22103683.

[6] A. C. Farjardo and Y.-J. Kim, "Development of Filipino Phonetically-balanced Words and phoneme-level Hmms," IJARCCE, pp. 1–6, Jan. 2015, doi: 10.17148/ijarcce.2015.4101.

[7] F. R. Jr. Arnel Fajardo, "Convolutional Neural Network for Automatic Speech Recognition of Filipino Language," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 1.1 S I, pp. 34–40, Feb. 2020, doi: 10.30534/ijatcse/2020/0791.12020.

[8] E. H. Liao, K. Ganareal, C. C. Paguia, C. Agreda, M. Octaviano, and R. Rodriguez, "Towards the Development of Automatic Speech Recognition for Bikol and Kapampangan," 2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), Nov. 2019, Published, doi: 10.1109/hnicem48295.2019.9072783.

[9] M. A. Ruiz and C. R. Mendoza, ""Using Hardware-based Voice Recognition toInteract with a Virtual Environment", Virtual Reality Laboratory, Proceedings of the IVEVA 2004 Workshop. Intelligent Virtual Environments and Virtual Agents University of Colima, CEUPROMED, Colima, 28040,Mexico.

[10] J. J. Raval and R. Gajjar, "Real-time Sign Language Recognition using Computer Vision," *2021 3rd International Conference on Signal Processing and Communication (ICPSC)*, May 2021, Published, doi: 10.1109/icspc51351.2021.9451709.

[11] K. Amrutha and P. Prabu, "ML Based Sign Language Recognition System," *2021 International Conference on Innovative Trends in Information Technology (ICITIIT)*, Feb. 2021, Published, doi: 10.1109/icitiit51526.2021.9399594.

[12] V. Adithya and R. Rajesh, "Convolutional Neural Network based Sign Language Recognition to Assist Online Assessment," *2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS)*, Sep. 2021, Published, doi: 10.1109/access51619.2021.9563346.

[13] W. Qin, X. Mei, Y. Chen, Q. Zhang, Y. Yao, and S. Hu, "Sign Language Recognition and Translation Method based on VTN," *2021 International Conference on Digital Society and Intelligent Systems (DSInS)*, Dec. 2021, Published, doi: 10.1109/dsins54396.2021.9670588.

[14] C. Suardi, A. N. Handayani, R. A. Asmara, A. P. Wibawa, L. N. Hayati, and H. Azis, "Design of Sign Language Recognition Using E-CNN," *2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT)*, Apr. 2021, Published, doi: 10.1109/eiconcit50028.2021.9431877.

[15] J. L. Bautista and Y. Kim, "[PDF] An Automatic Speech Recognition for the Filipino Language using the HTK System | Semantic Scholar," *[PDF] An Automatic Speech Recognition for the Filipino Language using the HTK System | Semantic Scholar*, Jan. 01, 2020. https://www.semanticscholar.org/paper/An-Automatic-Speech-Recognition-for-the-Filipino-Bautista-Kim/7d47daadfa9738461db5073afaa0f4798baed86e.

[16] A. F. Laguna and R. C. Guevara. Development, Implementation and Testing of Language Identification System for Seven Philippine Languages. *Philippine Journal of Science* 144 (1): 81-89, June 2015 ISSN 0031 – 7683.

[17] U. Fayyad, "Knowledge discovery in databases: An overview," Inductive Logic Programming, pp. 1–16, 1997, doi: 10.1007/3540635149_30.

[18] Frosch, Hannah N., "Spectral Analysis and Comparison of Analog and Digital Recordings" 2017. Honors Theses. 545. Retrieved from: https://aquila.usm.edu/honors_theses/545/.