Development of Speech to Filipino Sign Language for the Deaf or Mute Filipinos using Hidden Markov Model (HMM)

Raymond Joseph Meimban, Christian Lontoc Pao, Rogelio B Aniez jr., Julie Ann Loresco

Engineering Department, National University, Manila, Philippines.

Emails: rrmeimban@national-u.edu.ph; clpao@national-u.edu.ph; rjraniez@national-u.edu.ph; jahloresco@national-u.edu.ph

***Abstract* - The main purpose of this project is to design and develop a software based computer program that will convert Filipino Speech into its equivalent Sign Language. This program will run in the personal computer through MATLAB and will use Hidden Markov Model (HMM) as speech recognition engine that uses a MFCC coefficient in improving the speech feature representation in a HMM based recognition speech. It uses 7 state left-to-right model and a word-based model for a small scale vocabulary speech recognition. This study involves testing the system's speech recognition reliability and accuracy in which it is indicated in the objectives of the project. The analysis is expected to meet the 85% reliability of the speech recognition of the system proposed in the objective.**

Keyword - MATLAB, Speech to Sign Language, Accuracy, Speech recognition, HMM (Hidden Markov Model).

I. INTRODUCTION

Communication is the process of exchanging information by sending and receiving. Humans communicate with each other on a lot of different ways, from a simple talk personally, through computers, snail mail, cellular phones, internet, and many more communication gadgets. It is their way of spreading knowledge and information. Humans use the five senses, sight through the eye, hearing through the ear, taste through the tongue, touch through the skin, and smell through the nose in order for them to communicate. Moreover, communication is the foundation of all human relationship. But there are cases where these senses are lost or weakened, due to birth defect, from hereditary, accidents, or just by aging. Those people who are deaf from birth is considered as mute, because they cannot learn how to speak without being able to hear. On the other hand, hearing people can be mute if the physical apparatus of speech such as the vocal cords, tongue, and voice box is broken or damaged. This occurs less frequently than do speech problems related to hearing loss.

According to a conducted census that is subjected for the percentage/numbers of persons with disabilities of the National Statistical Office (NSO), an organization handled by the Philippine Government, there are only 942,098 persons with disabilities that lives in the Philippines, representing 1.2% of the total population or 352,398 of the total population have a disability in both hearing & speaking and 8.1% or 76,731 of the total population have a disability in hearing itself only. Sign language helps lessen the gap between those hearing people and those who are not. Using a collection of hand gestures and symbols, sign language is an effective means of communication that provides the deaf and those hard of hearing a way to interact with the world around them. But to be able to communicate with the deaf or mute, one must know how to use the Filipino sign language. This study is focused in the research and development to the basic Filipino words. Speech recognition is a vast field in our technology that will not be limited to just speech. Now, it will be extended for the aid of the deaf or mute Filipinos. This study is encouraging updates and modification due to the wide variations of language. Also, language has different ways of pronunciations depending on the user. It is also limited to the deaf and mute Filipinos because the study is focused to the sign language output.

Communication, the process of sending and receiving information, the person sending the message is referred to as the sender, while the person receiving the information is called the receiver. There are three basic categories of communication, the verbal, nonverbal, and written communication [8]. An example of nonverbal communication is sign language. Sign language is the way used in communicating with an individual who has an inability and incapacity in hearing who is called deaf or mute. However, not everyone knows how to use sign language.

The National Statistical Office (NSO), an organization handled by the Philippine Government, conducted a census that is subjected for the percentage/numbers of persons with disabilities. Base on the census, there are only 942,098 persons with disabilities that lives in the Philippines, representing 1.2% of the total population or 352,398 of the total population have a disability in both hearing & speaking and 8.1% or 76,731 of the total population have a disability in hearing itself only.

According to Department of Defense Hearing Center Excellence (HCE), it is said that hearing is the most essential survival sense, to which sound is mostly the source of information of the warriors against the enemies, which

makes hearing important for both lethality and survivability. In military, for a mission to be accomplished it is necessary to have teamwork, reporting skills and especially hearing as a fundamental way to understand the instructions. Not just in military but also in the real world, hearing is very essential to every human. People use their ears to build relationship and to have connections with friends and family, fully participating in team and community activities, and experiencing life events.

.

II. SYSTEM DESIGN OVERVIEW

A. System Algorthm

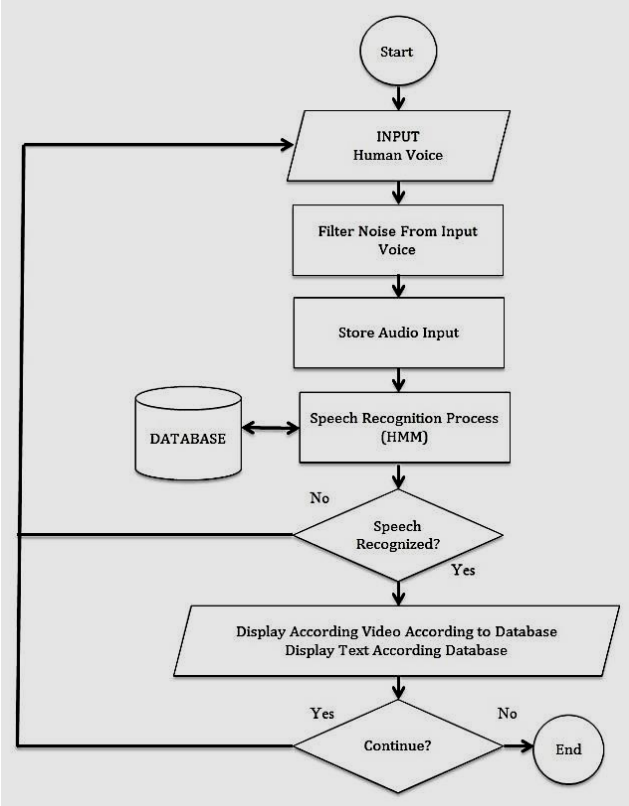

Fig. 1. System Algorithm

Figure shows the sequential process of the system. The first step is the execution of speech input where the user will speak through the microphone and the microphone will convert the voice input into electrical signal. After the speech enters the system, the system will filter the noise from the electrical. After the electrical signal is filtered from noise, it will be stored temporarily for processing and recognition. After it has been recognized and processed, the system will import its equivalent sign language video from the database. After the equivalent sign language video of the input speech is acquired, the speech will be identified by the system as recognized and the sign language video will be displayed in the GUI. After the execution of the sign language video, the user will decide if he wants to continue or end.

B. System Block Diagram

The figure 2 block diagram presents the sequential flow of operation of the system. The first step is the speech input through free space into the unidirectional microphone. The microphone converts the sound waves into an electrical signal. When the sound waves strike the microphone's diaphragm, they will cause it to move within an electro-magnetic field, which, in turn, creates a variance in an electrical current. The microphone is connected to the personal computer through a USB cable. The sound waves that had been converted into an electrical signal will now be transferred inside the personal computer. The MATLAB software will be loaded into the computer's storage then it will be executed. The electrical signal will be processed using the HMM. The HMM will be used to model a short time stationary signal series. A HMM for a sequence of words or phonemes is made by concatenating the individual trained HMM for the separate words and phonemes. After the electrical signal is processed and recognized, the system will look inside the database or the DAT file through the storage of the personal computer for the required file in order to obtain the equivalent sign language created in videos. After the equivalent sign language videos were identified, the MATLAB will import those sign language videos and display it videos on its self-contained GUI feature.

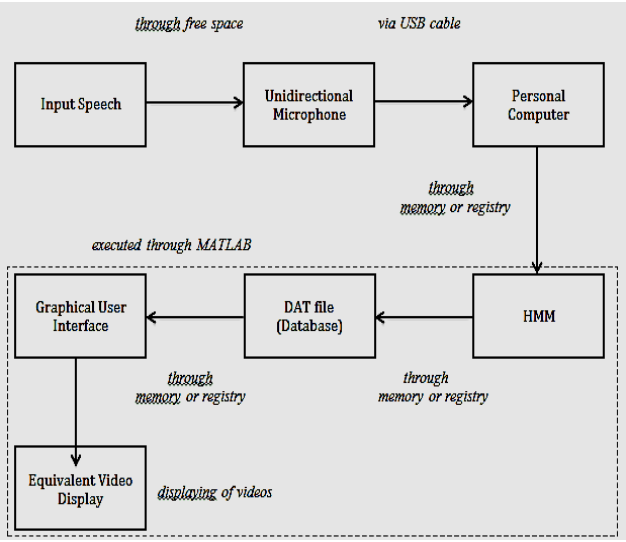

Fig. 2. System Block Diagram

III. METHODOLOGY

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note

sections A-D below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

A. Test of Accuracy

TABLE I. STATES TEST OF ACCURACY

Number of States	Accuracy (%)
3	80
4	83
5	84
6	86
7	89
8	88
9	85
10	84

In order to obtain the number of states that will be used, the researchers have conducted a test of accuracy for each state. This test was conducted using the whole vocabulary and it has a total of 5 test data for each vocabulary. Table I represents the test conducted in this project that determines the number of states that will fit in this system for its best performance. The first column represents the number of states and the 2nd column represents its equivalent accuracy in percentage. The data shows that the accuracy is increasing from state 3 up to the state 7, but after state 7, the accuracy starts to decrease. The researchers have concluded from this data, that 7 states is the number of states for optimum performance of this system.

B. Transition Probability Initialization

The transition guess is initialized because the training process requires an initial guess in order to have a reference for the estimated transition that will be generated at the output of the training process given that this system works on a 7 states left to right Hidden Markov Model.

It is necessary to initialize these model parameters, the most important initialization for a speech model is the transition matrix because it gives the system a chance to constrain the possible transitions.

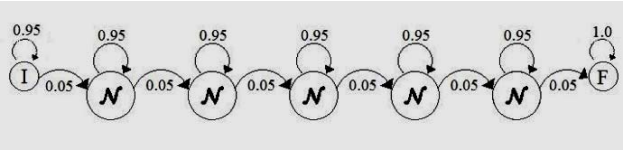

Fig. 3. Hidden Markov Model State

C. Testing Phase

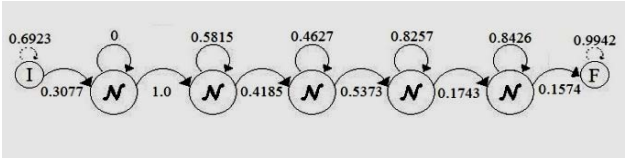

Fig. 4. Estimated Transition Model for the word "Disyembre"

TABLE II. ESTIMATED TRANSITION MATRIX FOR THE WORD "DISYEMBRE"

	State 1	State 2	State 3	State 4	State 5	State 6	State 7
State 1	0.6923	0.3077	0	0	0	0	0
State 2	0	0	1	0	0	0	0
State 3	0	0	0.581	0.418	0	0	0
State 4	0	0	0	0.462	0.5373	0	0
State 5	0	0	0	0	0.825	0.174	0
State 6	0	0	0	0	0	0.842	0.157
State 7	0	0	0	0	0	0	1

Figure 4 represents the 7 states left to right HMM estimated transition model of the word "Disyembre" and table II shows its equivalent form in matrix. According to these data, from the initial state or state 1, it has a 0.6923 probability to stay in its own state, and 0.3077 probability to move in the second state. State 2 has 0 probability to stay in its own state, and a probability of 1 to move on the 3rd state. State 3 has a 0.5815 probability to stay in its own state and 0.4185 probability to move in the 4th state. State 4 has a 0.4627 probability to stay in its own state and 0.5373 probability to move in the 5th State. State 5 has 0.8257 probability to stay in its own state and 0.1743 probability to move on the 6th state. State 6 has 0.8426 probability to stay in its own state and 0.1574 probability to move on the 7th State. State 7 is the final state.

Each state only has 2 transition paths. It has a probability to stay in the present state and a probability to move in the next state, which has a total probability of 1, except for the 7th or final state which has a 100 percent probability to stay at rest.

In this HMM, a single state doesn't necessarily represent a boundary or a single phoneme. In this whole word HMM, the speech may be segmented randomly. A single state may represent a word, phoneme, words and phoneme, a single letter, etc., and it doesn't have an exact boundary. In this example of transition model shown in figure 17, this model, represents the word "Disyembre" but it cannot be concluded each state in that model represents every phoneme of the word "Disyembre". Each and every state in this sample may be a part of that word without having an exact boundary.

IV. STATISTICAL ANALYSIS AND RESULTS

The suggested vocabulary, letters and phrases are grouped, and carefully organized into the table. To obtain

the number of tests that will be conducted for each data in order to compute the accuracy and reliability of the system, the proponents will be using Slovin's Formula.

Slovin's Formula:

$$n = \frac{N}{1 + Ne^2} \quad (1)$$

Where:
n= number of tests
N = total population
e= margin of error

With the total population of 700 persons, the researchers decided to have a confidence level of 90% for an entire population or a margin of error 10%. The total number of tests in order to have a more reliable result will be computed using Slovin's formula.

Therefore,

$$n = \frac{700}{1 + (700)(0.10)^2}$$

$$n = 88\ tests$$

A total number of 88 tests will be conducted to each of the lexicon's entry in order to compute for the accuracy and reliability of the system.

The overall accuracy of the speech recognition is 86.62%. It is computed by getting the average of all the entry's accuracy. The accuracy of each is computed by subtracting one from the total number of tests minus the sum of the tests of male and female divided by the total number of tests times 100%.

The reliability of the system is computed by dividing the total number of successful remarks to the sum of the total number of remarks. There are 134 entries in the lexicon, 116 indicates the total number of successful tests while 18 is the total number of unsuccessful tests. A successful remark means that the computed accuracy of that single entry is greater than or equal to 85%, while the unsuccessful remark means that the computed accuracy is below 85%.

$$\%R = \%R = \frac{116}{116 + 18} \times 100\% = 86.56\%$$

V. DISCUSION AND CONCLUSION

The group had successfully built a Filipino Speech to Sign Language using MATLAB based program as a workplace. The design recognizes comprehensible Filipino speech and transforms it into its equivalent sign language. This study's primary goal is to be an aid in communication with the deaf or mute individuals. It aims to help hearing impaired individuals with their insufficiencies to express their feelings to ordinary people, for them to survive with the advancement of technology that we have today.

The researchers were able to develop a speech recognition system that complies with an agreed recognition rate of 85% based on the multiple tests that has been made. To be able attain an 85% reliability; the group concluded that the number of recordings for training should be atleast 20 and higher. The more number of recordings per training the more accurate the system will be. To assess the performance of the system, a series of tests were conducted in My Children's House of Hope, Bahay Bata 127 and Payatas B Elementary School, consisting of 44 tests for each gender.

REFERENCES

[1] deaf philippines. (2008, September 30). Retrieved November 3, 2015, from deaf philippines word press: https://deafphilippines.wordpress.com/tag/philippine-school-for-the-deaf/

[2] is251Spring2010. (2010, February 3). Retrieved October 14, 2010, from is251Spring2010 Web site: http://is251spring2010.wikispaces.com/share/view/19721305

[3] Hope Speak. (2014, December 16). Retrieved September 28, 2015, from Hope Speak Web site: https://www.hopespeak.com/blog/why-is-communication-important-to-human-life-2/

[4] Blog Classroom Teachear. (2015, July 1). Retrieved October 14, 2015, from Blog Classroom Teachear Web site: http://blog.classroomteacher.ca/2576/compare-speakq-speech-assistive-technology-dragon-naturallyspeaking-voice-recognition-software/

[5] Wikipedia. (2016, October 5). Retrieved October 5, 2016, from Wikipedia: https://en.wikipedia.org/wiki/Baum%E2%80%93Welch_algorithm

[6] Bautista, J. L., & Kim, Y.-J. (n.d.). worldcomp-proceedings. Retrieved November 6, 2015, from worldcomp-proceedings web site: http://worldcomp-proceedings.com/proc/p2014/ICA3221.pdf

[7] Blanca, A. B. (2015, May 21). Sign Language. (L. E. Cuaresma, J. A. Erguin, A. J. Lucio, & E. Mendoza, Interviewers)

[8] Carpenter, M., Bauer, T., & Erdogan, B. (2015). flat world knowledge. Retrieved November 7, 2015, from flat world knowledge Web site: http://catalog.flatworldknowledge.com/bookhub/5?e=carpenter-ch12_s02

[9] Cassidy, S. (2002). Macquarie University. Retrieved November 6, 2015, from Macquarie University Web site: http://web.science.mq.edu.au/~cassidy/comp449/html/index.html

[10] Chua, J., Chua, U., De Padua, C., Tan, J., & Cheng, D. (2011). de Lasalle University. Retrieved January 23, 2015, from de Lasalle University Web site: http://www.dlsu.edu.ph/research/centers/adric/nlp/_archive/8th-NNLRS/8nnlprs-58.pdf

[11] Davis, S., & Mermelstein, P. (1980).

[12] deaf web sites. (n.d.). Retrieved September 28, 2015, from deaf web sites: http://www.deafwebsites.com/learning-sign-language.html

[13] El-Darymli, K., Khalifa, O., & & Enemosah, H. (n.d.). Academia. Retrieved January 23, 2015, from Academia Web site: http://www.academia.edu/1366077/Speech_to_Sign_Language_Interpreter_System_SSLIS_

[14] Frädrich, L., & Anastasiou, D. (2012, July). Translate Journal. Retrieved October 14, 2015, from Translate Journal Web site: http://translationjournal.net/journal/61dictating.htm

[15] Gales, M., & Young, S. (2008).

[16] Halawani, S., & Zaitun, A. (2012). Ripublication. Retrieved January 23, 2015, from Ripublication Web site: http://www.ripublication.com/ijise_info/ijisev2n1_02.pdf

[17] Hearing Center of Excellence. (n.d.). Retrieved October 7, 2015, from Hearing Center of Excellence Website: http://hearing.health.mil/PreventionProtection/ImportanceofHearing.aspx

[18] Huang, X. (n.d.). Association for Computational Linguistics. Retrieved November 6, 2015, from Association for Computational Linguistics: http://aclweb.org/anthology/H/H91/H91-1054.pdf
[19] Kile, L. (2015, May 11). Speech Technology. Retrieved October 14, 2015, from Speech Technology Web site: http://www.speechtechmag.com/Articles/News/Speech-Technology-News-Features/Sensory-Releases-TrulyNatural-Deep-Neural-Net-Speech-Recognition--103878.aspx
[20] Master, N. (2013, May 20). Rfgen. Retrieved November 6, 2015, from Rfgen web site: http://www.rfgen.com/blog/bid/292771/What-is-the-Difference-Between-Speaker-Dependent-Software-and-Speaker-Independent-Software
[21] Mathworks. (n.d.). Retrieved January 23, 2015, from Mathworks Web site: http://www.mathworks.com/products/matlab/index.html
[22] Mathworks. (n.d.). Retrieved January 23, 2015, from Mathworks Web site: http://www.mathworks.com/help/stats/hidden-markov-models-hmm.html
[23] Mathworks. (n.d.). Retrieved Ocotber 5, 2016, from Mathworks: http://www.mathworks.com/help/stats/hmmdecode.html
[24] Mathworks. (n.d.). Retrieved April 1, 2016, from Mathworks: https://www.mathworks.com/matlabcentral/fileexchange/37351-face-recognition-software
[25] Mathworks. (n.d.). Retrieved April 1, 2016, from Mathworks: https://www.mathworks.com/matlabcentral/fileexchange/32849-htk-mfcc-matlab/content/mfcc/mfcc.m?requestedDomain=www.mathworks.com&requestedDomain=www.mathworks.com
[26] Media College. (n.d.). Retrieved April 9, 2015, from Media College Website: http://www.mediacollege.com/audio/microphones/directional-characteristics.html
[27] Microsoft Developer Network. (n.d.). Retrieved October 16, 2015, from Microsoft Developer Network web site: https://msdn.microsoft.com/en-us/library/hh378451(v=office.14).aspx
[28] Novet, J. (2015, May 28). Venture Beat. Retrieved October 14, 2015, from Venture Beat website: http://venturebeat.com/2015/05/28/google-says-its-speech-recognition-technology-now-has-only-an-8-word-error-rate/
[29] Paul, D. (1990). Massachusets Institute of Tecgnology. Retrieved January 23, 2015, from Massachusets Institute of Tecgnology Web site: http://www.ll.mit.edu/publications/journal/pdf/vol03_no1/3.1.3.speechrecognition.pdf
[30] Philippine Statistic Authority. (n.d.). Retrieved September 28, 2015, from Philippine Statistic Authority Web site: https://psa.gov.ph/content/persons-disability-comprised-123-percent-total-population
[31] Philippines Statistic Authority. (n.d.). Retrieved October 7, 2015, from Philippines Statistic Authority Web site: http://senate.gov.ph/lisdata/77546232!.pdf
[32] Ramanathan, N. (2006). University of Maryland. Retrieved January 23, 2015, from University of Maryland Web site: http://www.cs.umd.edu/~djacobs/CMSC828/ApplicationsHMMs.pdf
[33] skwirk. (n.d.). Retrieved September 28, 2015, from skwirk Web site: http://www.skwirk.com/p-c_s-11_u-26_t-74_c-265/five-senses/nsw/five-senses/way-out-communication/human-communication
[34] Uchat, N. S. (n.d.). Department of Computer Science and Engineering. Retrieved September 15, 2015, from Department of Computer Science and Engineering Web site: https://www.cse.iitb.ac.in/~nirav06/i/HMM_Report.pdf
[35] Young, S., Odell, J., Ollason, D., Vatchev, V., & P, W. (1997). The HTK Book. Cambridge: Cambridge University Press.