English-Cebuano Word Pronunciation Checker using Simplified Linear Predictive Coding

A Thesis

Presented to

the Faculty of the Department of Computer Science

University of San Carlos

In Partial Fulfillment

of the Requirements for the Degree of

Bachelor of Science in Computer Science

by

Igot, Miraflor C.

Yap, B. Carla O.

Jacqueline F. Yara

Faculty Adviser

October 2012

**Abstract**

“Mother Tongue-Based Multi-Lingual Education” (MTB-MLE) is part of the K+12 basic education reform program and the Cebuano language is part of the languages being identified by the Department of Education. The lack of materials needed to teach the students in reading, writing and speaking Cebuano became an issue. In addition, the foreigners staying in Cebu are trying to learn to speak the language.

A pronunciation checker helps improve reading, writing, and speaking ability of the people on the language of Cebuano. A pre-recorded audio file will guide the person on speaking the proper pronunciation once he has searched for the English word’s translation. The person speaks through the microphone and records his voice with a sampling rate of 44100 kHz. His voice in the audio file will undergo Silence Removal (SR). Linear Predictive Coding (LPC) is used to extract the speech content in the order of 10. To check the similarities and closeness of the pronunciation, Dynamic Time Warping (DTW) is utilized to stretch or compress the speech content.

The algorithm works of about 68% in terms of accuracy categorized by a three-point interval rating scale. However, 90% is estimated to have obtained the correct pronunciation. The results that the algorithm from this study gives denote accurate values that are significant enough in identifying the wrong pronunciation from the right one.

**Table of Contents**

[Table of Contents 1](#_Toc338852002)

[CHAPTER 1 - INTRODUCTION 1](#_Toc338852004)

[1.1 Rationale of the Study 1](#_Toc338852005)

[1.2 Research Objectives 2](#_Toc338852006)

[1.2.1 General Objectives 2](#_Toc338852007)

[1.2.2 Specific Objectives 2](#_Toc338852008)

[1.3 Scope and Limitations of the Research 2](#_Toc338852009)

[1.4 Significance of the Research 3](#_Toc338852010)

[1.5 Research Methodology 3](#_Toc338852011)

[CHAPTER 2 - REVIEW OF RELATED LITERATURE 5](#_Toc338852013)

[CHAPTER 3 - THEORETICAL FRAMEWORK 12](#_Toc338852015)

[3.1 Pronunciation 12](#_Toc338852016)

[3.2 Cebuano Language 12](#_Toc338852017)

[3.3 Speech Signals 13](#_Toc338852018)

[3.4 Linear Predictive Analysis 15](#_Toc338852019)

[3.5 Dynamic Time Warping 17](#_Toc338852020)

[CHAPTER 4 - THE ALGORITHMS AND THEIR MODELS 20](#_Toc338852022)

[4.1 Silence Removal and Noise Reduction 22](#_Toc338852023)

[4.2 Frame Blocking and Hamming Window 25](#_Toc338852024)

[4.3 Feature Extraction – Linear Predictive Coding 26](#_Toc338852025)

[4.4 Pattern Matching – Dynamic Time Warping 27](#_Toc338852026)

[4.5 Data Source 28](#_Toc338852027)

[4.6 Database Model 30](#_Toc338852028)

[4.7 Definition of Terms 30](#_Toc338852029)

[CHAPTER 5 - RESULTS AND ANALYSIS 32](#_Toc338852031)

[5.1 Softwares Used 32](#_Toc338852032)

[5.2 Determining Sample Size 36](#_Toc338852033)

[5.3 Test Results 37](#_Toc338852034)

[5.4 Evaluation 47](#_Toc338852035)

[CHAPTER 6 - CONCLUSION AND RECOMMENDATIONS 52](#_Toc338852037)

[6.1 Conclusions 52](#_Toc338852038)

[6.2 Recommendations 52](#_Toc338852039)

[APPENDIX A Curriculum Vitae 53](#_Toc338852040)

[APPENDIX B Consultation Details 55](#_Toc338852041)

[References: 57](#_Toc338852042)

**List of Tables**

[Table 2.1 Mnemonic Values of the Cebuano Consonants 4](#_Toc338588313)

[Table 2.2 Mnemonic Values of the Cebuano Vowels 5](#_Toc338588314)

[Table 3.5.1 Signal Difference 17](#_Toc338588315)

[Table 4.5.1 100 Commonest English Words 26](#_Toc338588316)

[Table 5.3.1 Test Results: Adult Female 36](#_Toc338588317)

**List of Figures**

[Figure 2.1 Frequency Wave 7](#_Toc337497588)

[Figure 2.2 (a) sample analog signal; (b) digital signal after conversion 8](#_Toc337497589)

[Figure 2.3 PCM Encoded Signal 9](#_Toc337497590)

[Figure 3.3.1 Frame-by-Frame Processing 13](#_Toc337497591)

[Figure 3.5.1 Test Pattern and Reference Pattern 16](#_Toc337497592)

[Figure 4.1 General Architecture 18](file:///C:\Users\carla\Dropbox\thesis\commented%20body%20file.docx#_Toc337497593)

[Figure 4.1.1 Raw file of ‘maayo’ 20](#_Toc337497594)

[Figure 4.1.2 ‘maayo’ file after silence removal 21](#_Toc337497595)

[Figure 4.2.1 One Frame Block of ‘maayo’ file 22](#_Toc337497596)

[Figure 4.2.2 ‘maayo’ signal after windowing 23](#_Toc337497597)

[Figure 5.1 Use Case Diagram 28](#_Toc337497598)

[Figure 5.2 Control Flow Diagram 29](#_Toc337497599)

[Figure 5.3 Class Diagram 30](#_Toc337497600)

[Figure 5.4.1 Results 36](#_Toc337497601)

[Figure 5.4.2 Cause of the Errors 37](#_Toc337497602)

[Figure 5.4.3 Evaluation of Pronunciation Mean Scores 38](#_Toc337497603)

# CHAPTER 1

# INTRODUCTION

## 1.1 Rationale of the Study

There has been a rise of the number of foreigners here in the Philippines. The slogan “It’s more fun in the Philippines” has finally worked, as Iain Jamieson—Visa country manager for the Philippines and Guam—reported the sudden double-digit level growth in inbound spending by foreigners hitting 12-13 percent in November 2011 (Dumlao, 2012). Korean and Singaporean tourists in Cebu are growing, 47.82% and 38.54% respectively (Aznar, 2012). It is also not surprising that the Cebuano traders will be doing business with Australia and New Zealand that a conference is held at the Parklane International Hotel on March 23, 2012 (Mozo, 2012).

As Cebu is promoting tourism and business, it is also possible to support the learning of the Cebuano or Visayan language. It is the second most-widely spoken language in the country. Even Globe Telecom has provided a Visayan customer service, to make “Globe experience more personal”, as was said in The Philippine Star, February of 2012. The Cebu Provincial Board (PB) used the vernacular as medium of communication in their session on February 27, 2012, and it was embarrassing to know that some of the PB members stammered in pronouncing Cebuano words like ‘komitiba’ (committee) and ‘daklit’ (to narrate) (Adlawan, 2012). More recently, Cebuanos were taken to Vladivostok, Russia to man front reception desks and supervise hotel staff services for the APEC leaders’ summit held on September 6, 2012. In effect, the Russian counterparts have tried to master some Cebuano words (Jumilla, 2012).

The Department of Education (DepEd) further helps the case especially with the PB indirectly, as DepEd pushes the “Mother Tongue-Based Multi-Lingual Education” (MTB-MLE) as part of the K+12 basic education reform program in all public schools, starting at June 2012. Based on DepEd’s Order No. 16 regarding the implementation of the MTB-MLE, the student’s mother tongue will be used to teach in all learning areas from kindergarten to Grade 3, except in Filipino and English subjects. The program also hopes to focus on academic development which prepares the learner to acquire mastery of language and culture and socio-cultural awareness (Yu, 2012).

It is noticeable that promoting the Cebuano language has become academic in nature as well. The field of Computer Science can connect to it and become part of the program, because technology is growing and is prevalent in the country.

## 1.2 Research Objectives

### 1.2.1 General Objectives

To create an English-Cebuano Word Pronunciation Checker that will provide a Cebuano equivalent of the inputted English word (in text). It is accompanied by an audio file with the proper pronunciation of that Cebuano word.

### 1.2.2 Specific Objectives

Specifically, the study intends to:

1. build an English word lookup with the Cebuano equivalent of each word;
2. record audio files as basis for the correct pronunciation of the word. One for the male and another for the female;
3. check user’s pronunciation by developing a Voice Comparison feature algorithm according to similarity;
4. test the algorithm; and
5. evaluate the created algorithm.

## 1.3 Scope and Limitations of the Research

The look-up table will focus on the vocabulary, and not the meaning of the word since it will be assumed that the user will be searching for the translation of the word, and not the definition. An input of the English language is mandatory and an existing English-Cebuano translator found at a website: cebu.sandayong.com will be used. The dialect to base on will be the language used by the Sugboanons (Cebuanos); the people residing in Cebu. There will be two sample audio files; one with a female voice and the other for a male voice. The models used for the audio files will be Sugboanons. Random groups of people will be asked for testing by recording their voices and further analyzing results with the use of the algorithm.

## 1.4 Significance of the Research

People will be able to learn the Cebuano Language with the proper pronunciation, without the need of buying a dictionary and additional reading of the phonemes since an audio file will be provided and can be followed. Teachers under the MTB-MLE can use this to base the words they will use for the class. Students as well can use this when doing Cebuano essays for their homework. The application will promote learning of the Cebuano language as well as the proper pronunciation of the words.

## 1.5 Research Methodology

*Phase I: Data Gathering*

Recordings of both male and female voices served as a listening guide to the correct pronunciation of the words. The database was created with the use of XAMPP. The most common English words were placed with their equivalent Cebuano translation and their form; adjective, noun or verb. The audio files are not however found in the database. The Cebuano translation was an acting ID to grab the path of the audio file, which was essentially inside the Net Beans project.

*Phase II: Designing and Development of the Algorithm*

By utilizing the different capabilities of algorithms designed for voice comparison feature, a more simplified algorithm was generated and used for the implementation of the study.

*Phase III: Testing*

The algorithm was tested in a silent room; male and female. A microphone attached to the laptop was used to capture the voice input. The inputs were correctly pronounced words and intentionally mispronounced words. Once captured and recorded, the algorithm takes place and displays the result.

*Phase IV: Evaluation*

The result of the algorithm were evaluated using a study by Eamonn Keogh and Michael Pazzani (2001). The study dealt with the classic Dynamic Time Warping that showcased a warped path of the similarities of two sequences or time series. The smaller the warping path, the closer are the similarities of the sequences.

# Chapter 2

# REVIEW OF RELATED LITERATURE

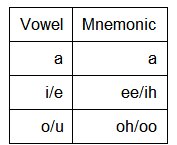
A phoneme is a representation of a sound. The Cebuano language consists of 16 phonemic consonants—sounds that each consonant represent—and three phonemic vowels (Endriga, 2010). The glottal stop ? is only spoken or pronounced and is not really seen in writing as shown in Table 2.1 and Table 2.2. The glottal stop is the sound similar to the middle of the English expression uh-oh (Räisänen et. al, 2007). It is a conventional consonant onset of orthographically vowel-initial words. It happens most often in the middle of the word between vowel sequences (Endriga, 2010). For example, the word ‘*maayo*’, which means ‘good’. A glottal stop is found after the first ‘a’ or before the second ‘a’. The glottal stop in a form of a hyphen mark occurs after a consonant like ‘*hilam-os*’ meaning ‘to wash one’s face’(Räisänen et. al, 2007). A glottal stop can also be found after the word itself. This holds true to the word ‘*basa*’ or ‘wet’. If the glottal stop is removed, ‘*basa*’ will be equivalent to the English word ‘read’.

Table 2.1 Mnemonic Values of the Cebuano Consonants

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Consonant | Mnemonic | Consonant | Mnemonic | Consonant | Mnemonic |
| B | Ba | L | La | R | Ra |
| K | Ka | M | Ma | S | Sa |
| D | Da | N | Na | T | Ta |
| G | Ga | NG | NGa | W | wa |
| H | Ha | P | P | Y | Ya |

Table 2.1 is the list of consonants with their respective sound equivalence, the mnemonic. Their mnemonics accompanies the vowel *a* as convention needed (Marking, 2005).

Table 2.2 Mnemonic Values of the Cebuano Vowels



The Table 2.2. is the list of vowels with their corresponding mneomonics. The ‘i’ and ‘e’ belong to one row; ‘o’ and ‘u’ go together as well.

The Cebuano language contains a fewer number of consonants compared to the English language. Some of these missing phonemes are being replaced with any of the phonemes found in Table 2.1, and is said to be of Spanish influence. The ‘f’ is replaced with the consonant ‘p’ as in the word ‘family’ turning into ‘*pamilya*’; the hard pronunciation of ‘c’ is replaced with ‘k’ like ‘carabao’ into ‘*kabaw*’; soft ‘c’ is represented as ‘s’ similar to the Spanish word ‘centavo’ or English word ‘cent’ which then becomes ‘*sintabo*’ for the Cebuano equivalent. The letter ‘q’ is replaced with two combined phonemes; ‘k’ and ‘w’. This applies to the word ‘*kwarta*’ or ‘money’. The ‘v’ is turned into ‘b’; ‘vacation’ becomes ‘*bakasyon*’ and the Spanish word ‘Noviembre’ or ‘November’ in English, translates to ‘Nobyembre’ in Cebuano. Some pronouns in the Spanish language like ‘Jose’ and ‘Juan’ retains the ‘h’ sound of the letter ‘j’ (Räisänen, et al., 2007).

In Table 2.2, the vowels provided are only three. The letter i and e are interchangeable as they are pronounced rather similarly in everyday speech. This pronunciation ruling also applies to the o and the u. An example word is the Cebuano equivalent of ‘big’ which is ‘*dako*’ or can be spelled as ‘*daku*’ although they sound just the same (Räisänen, et al., 2007). The ‘a’ is always pronounced like the first vowel of the English word ‘father’ (Marking, 2005).

The Cebuano is divided in two main dialects: the Cebu City dialect and the Provincial dialect (Räisänen, et al., 2007). In the Cebu City dialect, L is pronounced as W when it is between the following vowels: o/u and a; a and o/u. The word ‘*ulan*’ as in ‘rain’ becomes ‘*uwan*’; ‘*balod*’ meaning ‘waves’ becomes ‘*bawod*’. L is dropped when it is found to be in the middle of two a’s like the word ‘*halang*’ or ‘spicy’ which then becomes ‘*hang*’. For the Provincial dialect, L retains it sound (Marking, 2005).

A sampled sound of speech with isolated Cebuano words were recorded in the UCLA Phonetics lab on 1987 but was recently added to a digital database on August, 2008. The recording is available on two formats: the wav file and the mp3 file. These audio format are the most widely used (Gupta, 2009). The Department of Linguistics of The University of Pennsylvania points out that the best audio format for research purposes is the one that minimizes distortion. MP3 should not be chosen as it is lossy and distort signals for compression. It is optimized for music at higher sample rates than generally needed for speech (Frisbie, 2005).

WAVE or WAV which means ‘waveform’ is usually chosen for speech files. The AT&T Labs uses WAV for their speech output audio format. In a study of Speaker Verification, WAV file is also used to be stored in a database (Dumitru, et al., 2006). The WAV file format is said to be advantageous because they are lossless files; They do not lose the quality of the speech once stored. They can be edited or they can be manipulated relatively easily. (Paulus, 2008).

Ganesh Tiwari and his co-researchers explain the generation of speech by the lungs and diaphragm where the air passes through the larynx tube modulated by the vocal cords in the process called phonation or voicing. Phonation is responsible for the generation of pitch and tone. Articulation is followed, as the modulated air is filtered by the mouth, nose and throat. The emitted pressure wave excites the air and creates sounds, creating peaks of frequencies called formants. Frequency describes the number of waves that pass a fixed place in a given amount of time. Figure 2.1 shows a closely compressed wave as a high frequency and the wider wave as low frequency. These frequencies help discriminate the sounds from each other. When the frequency of a sound increases, the sound gets high-pitched and irritating. When the frequency of a sound decreases, the sound gets deepen (Çelik, 2008). Phonemes can then be characterized by the changing frequencies of formants (Tiwari, et al., 2011).

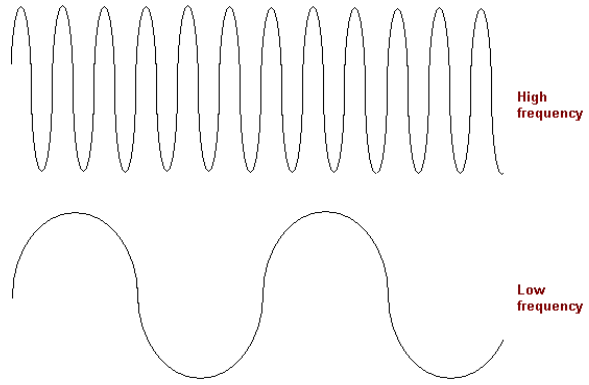


Figure 2.1 Frequency Wave

Figure 2.1 shows the differences between the low frequency and the high frequency. A higher frequency has a narrow closeness of curves, meaning there are many peaks found in the signal. The low frequency is the opposite; wider curves and infrequent.

Recording speech does not limit in only one gender. Due to the differences in vocal tract length, male, female, and children’s speech are different (Çelik, 2008). On research, it is often that participants are a mixture of male and female. In a study of recognition of Tagalog Alphabets, Roland Navarro Jr. used 25 males and 25 females. The Filipino Speech Corpus has recorded Filipino speech from 50 female and 50 male speakers. The research on Acoustic Characteristics on Vowel Space from the Department of Linguistics of the University of the Philippines Diliman also involves one speaker of each gender of each native Filipino. In Dumaguete, Costanilla and Montenegro of Silliman University supplies 50 different speakers with a mixture of male and female genders. The adult female vocal tract is on average 5.5 inches whereas the average male vocal tract is 6.6 inches. These physical differences between the sexes cause a difference in the way they speak. Not only do the pitch and frequency differ, there are also differences in word-forming patterns such as articulation and speed. Female vocal folds are shorter, lighter and vibrate at approximately 220 Hz -- twice that of the average male (Joseph, 2009).

Speech recognition is the process of extraction of linguistic information from speech signals. The linguistic information which is the most important element of speech is called phonetic information (Çelik, 2008). Ganesh Tiwari and his partners describe it much simpler as converting audio signal into words. The applications include voice commands, data entry, automating the telephone operator’s job in telephony, and many more. But, the work principle of speech recognition systems is roughly based on the comparison of input data to prerecorded patterns. (Çelik, 2008). In a broader sense, it can recognize speech without being targeted at single speaker (Wadhani, et al., 2011).

When the input data to an algorithm are too large to be processed and it is suspected to be highly redundant (much data, but not much information) then the input data will be transformed to a reduced representative set of features (also named features vector). Transforming the input data into the set of features is called features extraction (Gupta, 2009). In the course of speech recognition, feature extraction is a one important part (Costanilla, et al., 2011). It is an information retrieval method on audio signals. It also helps knowing the category of that information or features in a recording, which are silent, noise and the actual word. There are number of models used to extract features and among them are the Linear Predictive Coding (LPC) and the Mel Frequency Cepstrum Coefficients (MFCC). LPC is based on assumption that a speech sample can be approximated by a linearly weighted summation of determined number of preceding samples. The most popular feature extraction method, MFCC mimics the human hearing behavior by emphasizing lower frequencies and penalizing higher frequencies (Tiwari, et al., 2011).

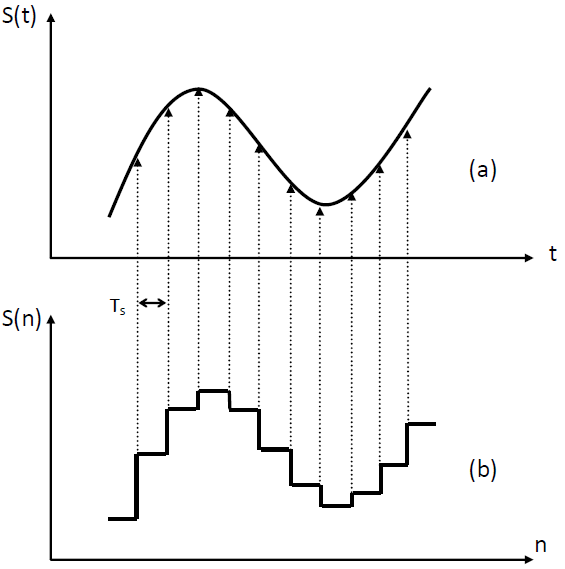


Figure 2.2 (a) sample analog signal; (b) digital signal after conversion

Figure 2.2 displays the transformation of the analog signal to digital. The digital signal creates a staircase like model, with the outer points as basis for the points in the analog signal.

Navarro (2007) explains that before any extraction procedure is done, the recording that records a speech signal must convert it to digital for easy processing. In the book of Digital Signal Processing by Proakis and Manolakis, conversion to digital signal is to convert analog signal into a sequence of numbers having finite precision, since the raw analog signal is a continuous time signal. Figure 2.2 (a) illustrates this conversion. But before getting a conversion of Figure 2.2 (b), the analog signal will undergo a pulse code modulation (pcm) illustrated on Figure 2.3. PCM is a technique used for telephone communication system (and also on Blu-ray and DVD) to digitalize an analog signal by taking a sample of a signal at uniform intervals and convert them to binary numbers.

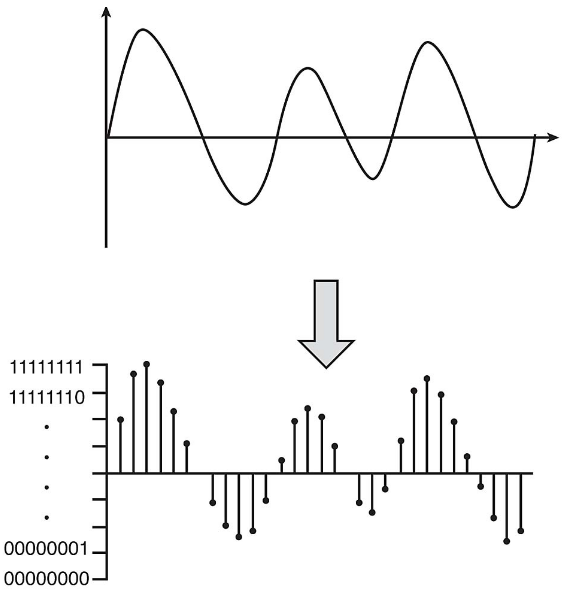


Figure 2.3 PCM Encoded Signal

The Figure 2.3 shows how the continuous signal is converted, before it can become the real digital signal. The PCM encoded signal is the bases to create the digital signal.

Pattern recognition is the most important component of the system (Wadhani, et al., 2011) as it is used to compare speech pattern in order to determine similarities (Costanilla et al., 2011). Pattern recognition holds different names but of the same meaning. These names include pattern matching, feature matching, decision making, and template matching. It is a branch of artificial intelligence inclined to distinguish patterns of interest from their background, and make sound and reasonable decisions about the categories of the patterns (Tiwari, et. al, 2011). Models for pattern recognition include Hidden Markov Model, Dynamic Time Warping, and Vector Quantization.

Navarro’s study (2007) on the recognition of Tagalog Alphabets makes use of the Hidden Markov Model (HMM), a finite automaton. It involves different observable states -- phonemes or characters in an alphabet -- over a given time but these states are hidden. They are hidden due to the reason that a word detected on a speech signal can have different states and the sequences of these states are too many and of different variations. Lawrence Rabiner explains the process further as an urn problem, a mental exercise on the subject of probability and statistics. There are three problems on HMM: during evaluation, does the sequence follow an existing model? What is the correct state sequence? How to create an optimized model to describe the sequence? (Dumitru, et al., 2006).

Vector Quantization (VQ) is a compression technique that selects the more effective features instead of using the whole. Rahul Mandal and Pawar Kumar Jha use VQ to create a more hybrid concept for text independent identification of a speaker. Speaker models are formed by clustering the speaker’s feature vector into a known cluster numbers. Each cluster is represented by a code vector—representative vectors—and the whole code vectors were saved at a codebook. In testing phase, the Euclidean metric was calculated for an unknown recorded signal and will be compared with the codebook of each speaker in the speaker database considering an acceptable distortion.

Stan Salvador and Phillip Chan (2007) explains dynamic time warping (DTW) as a technique that finds an optimal alignment between two time series in which one time series may be “warped” non-linearly by stretching or shrinking (compressing) its time axis locally. The distance between the two is computed, by summing the distances of individual aligned elements. This alignment can be used to find corresponding regions or to determine the similarity between the two time series.

# CHAPTER 3

# THEORETICAL FRAMEWORK

## 3.1 Pronunciation

It is possible for people to produce practically all the correct sounds but still be unable to communicate their ideas appropriately and effectively. On the other hand, people can make numerous errors in both phonology and syntax and yet succeed in expressing themselves clearly (Heaton, 1988). Pronunciation has been tested globally in different types of conversational exchange, ranging from an interview, reading aloud, or reports and presentations that go on inside and outside an educational institution (Bobda, 2011). The insufficiency of the testing of accuracy has hindered the assessment of a learner’s management of specific pronunciation features. This insufficiency is due to the particular difficulties involved in testing oral skills, such as administration and the lack of testing equipment, like tape recorders or laboratories.

Speech enables students to make connections between what they know and what they are learning, and listening helps them to acquire knowledge and explore idea. A successful speaker has some characteristics like highly motivated, a great degree of aptitude, flexible and adapt his/her style and beliefs to meet the demands of the task or the learning situation and also aware of importance of vocabulary, grammar and pronunciation. Moreover, he/she has complementary receptive skills (reading and listening) as models for what he/she wants to say, and she/he uses writing as a means of preparing for speaking encounters. Hence, a number of things are to be needed for fostering speaking skill (Narayanan, et al., 2008).

## 3.2 Cebuano Language

Cebuano is an Austronesian language spoken in the Philippines by about 20 million people, that is about 25% of the population. It is the most widely spoken member of the Visayan languages; a language closely related to Bahasa Melayu, Bahasa Indonesia, and other Filipino languages. It has the largest native language speaking population of the Philippines despite not being taught formally in schools and universities. It is the lingua franca of the Central Visayas and Mindanao (Cebuano language – Wikipedia, the free encyclopedia, 2012). Cebuano is a dialect that has evolved in different cities of the Philippines, especially in the Southern region. Massive migration has spread the language throughout the country, and today, about a quarter of the country’s population consider Cebuano as their native language (Cebu and Central Visayas – SE Asian Languages and Culture). Linguistically, it is the country’s second most widely used language (Learn Cebuano – Cebuano 101).

## 3.3 Speech Signals

The analysis of speech sounds takes into consideration their method of production. Under the speech analysis is a feature that highlights the formulation of a feature vector representation that captures the important information in the speech signal that will be used for future processing.

The speech signal is a slowly time varying signal station with its characteristics fairly stationary if examined between 5 and 100 msec. But on the order of 1/5 seconds or more, the signal characteristics change to reflect the different speech sounds being spoken. Usually first 200 msec or more (1600 samples if the sampling rate is 8000 samples/sec) of a speech recording corresponds to silence (or background noise) because the speaker takes some time to read when recording starts.

Pre-Processing of Speech Signal is very crucial in the applications where silence or background noise is completely undesirable (Saha, et al., 2005). Only the important data of the recorded signal is processed and that is the voiced part. The rest of the signal does not contain any required information. And also any silence part at the beginning or at the end of the signal will cause recognition failures. So finding the correct points for the beginning and end of the signal is very important (Çelik, 2008).

The captured audio signal may contain silence at different positions such as beginning of signal, in between the words of a sentence, end of signal, etc. If silent frames are included, modeling resources are spent on parts of the signal which do not contribute to the identification. The silence present must be removed before further processing (Tiwari, et. al., 2011).

All sound recordings can suffer from faults, though modern recording practice minimizes most of them to the point where they aren't a problem. However in the restoration of old (mostly [analogue](http://en.wikipedia.org/wiki/Analog_signal)) recordings there are a number of faults which are normal in recordings: the restoration process involves trying to eliminate or at least reduce them (Wilmut, R., 2011).

Noise is the main problem of sound recording, and over the years various methods have been applied to combat it. Noise usually takes the form of hiss, hum, whistles, and impulsive noises such as clicks and cackles.  It can be dealt with to an extent by simply filtering off the low frequencies in cases of low frequency continuous tones like hum, or by a 'low-pass' filter where it passes lower frequencies and blocks higher ones (Wilmut, R., 2011).

Because of the slowly varying nature of the speech signal, it is common to process speech in blocks (also called frames). The properties of the speech waveform can be assumed to remain relatively constant over very short intervals (Introduction to Speech Processing – OHSU, 2010).

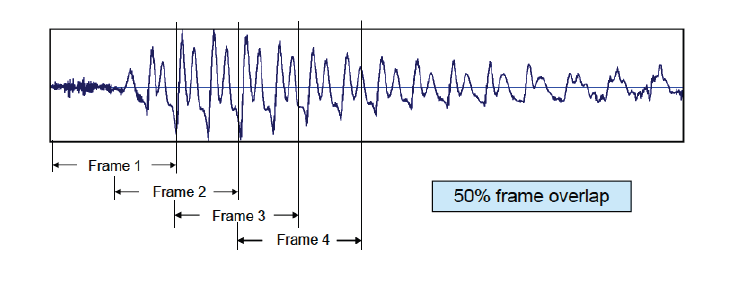


Figure 3.3.1 Frame-by-Frame Processing

Speech is processed frame-by-frame in overlapping intervals until entire region of speech is covered by at least one such frame as in Figure 3.3.1.

## 3.4 Linear Predictive Analysis

One of the most powerful speech analysis techniques is the method of linear predictive analysis. This method has become the predominant technique for estimating the basic speech parameters such as pitch, formants, vocal tract area functions, and for representing speech for low bit rate transmission (Linear predictive coding – Wikipedia, the free encyclopedia, 2012).

The basic idea behind linear predictive coding is that a sample of speech can be approximated as a linear combination of the past speech samples. By minimizing the sum of the squared differences (over a finite interval) between the actual speech samples and the linearly predicted ones, a unique set of predictor coefficients can be determined (Waibel, et. al., 1990). Linear Predictive Coding is often referred to as “inverse filtering”, as its aim is to determine the “all zero” filter. It is expected that the prediction error will be large at the beginning of each pitch period. Except for a sample at the beginning of each pitch period, every sample of the voiced speech waveform can be predicted from the past samples.

Since speech signals vary with time, the process of linear predictive coding is done on short chunks of the speech signal, which are called frames; generally 30 to 50 frames per second give intelligible speech with good compression. ((Linear predictive coding – Wikipedia, the free encyclopedia, 2012).

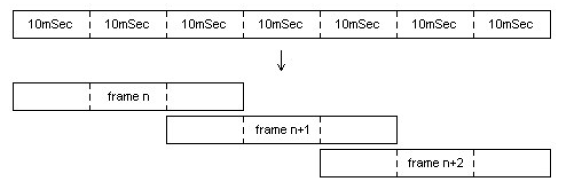


Figure 3.4.1 Audio Signal to Separate Frames

After the frames have been separated as seen in Figure 3.4.1, the LPC function will take every frame and extract the necessary information from it. This is the voiced/unvoiced, gain, pitch, and filter coefficients information. To determine if the frame is voiced or unvoiced you need to find out if the frame has a dominant frequency. If it does, the frame is voiced. If there is no dominant frequency the frame is unvoiced. If the frame is voiced you can find the pitch. The pitch of an unvoiced frame is simply 0. The pitch of a voiced frame is in fact the dominant frequency in that frame. One way of finding the pitch is to cross correlate the frame. This will strengthen the dominant frequency components and cancel out most of the weaker ones (ECE 352: Signals and Systems II – Oregon State University).

## 3.5 Dynamic Time Warping

Another established layer in speech processing is the stretching or compression of two time series in order to make one resemble the other as much as possible. The technique is called Dynamic time warping (DTW). It is a technique that finds an optimal alignment between two time series in which one time series may be “warped” non-linearly by stretching or shrinking its time axis. The alignment can be used to find corresponding regions or to determine the similarity between the two time series (Salvador, et. al., 2007). The output of the dynamic time warping algorithm is the remaining cumulative distance between the two time series.

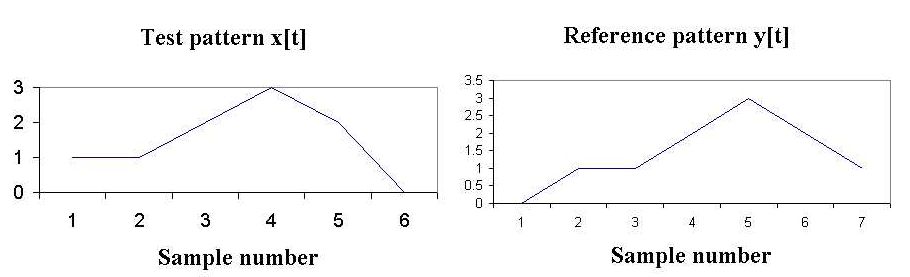


Figure 3.5.1 Test Pattern and Reference Pattern

Table 3.5.1 Signal Difference

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| a) (Input) test signal, *x*[*t*]: | 1 | 1 | 2 | 3 | 2 | 0 |  |
| b) (Stored) reference signal, *y*[*t* ]: | 0 | 1 | 1 | 2 | 3 | 2 | 1 |
| Sample-by-sample difference *x*[*t*] - *y*[*t*]: | 1 | 0 | 1 | 1 | -1 | -2 | undefined |

Above are two signals, a test signal and a reference signal. Both signals are similar in that they are single-peaked. However, the stored reference signal is longer than the test signal, and the peak is later. In other words, the two signals are not synchronized in time. In calculating the difference between the two signals, consider a matrix of the distance between every sample of x[t] and each sample of y[t].

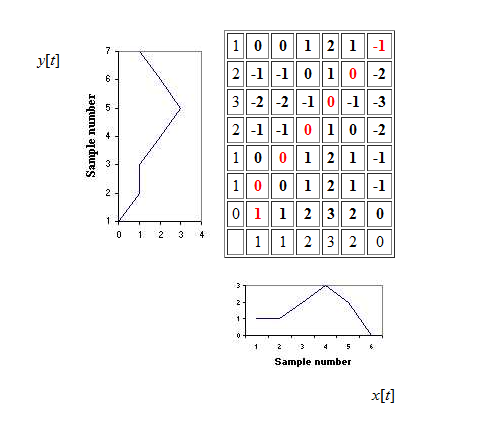


Figure 3.5.2 Distance Matrix D

There is a sequence of low numbers, close to the diagonal, indicating which samples of *x*[*t*] are closest in value to those of *y*[*t*] (Coleman, 2011). Aside from a simple subtraction, a symmetrical distance measure can also be used.

# CHAPTER 4

# THE ALGORITHMS AND THEIR MODELS

This chapter discusses the different approaches and the extracted capabilities of the algorithms incorporated in this research that will help in the full understanding of the study.

Figure 4.1 describes the overall structure and architecture of the system. It first starts on capturing the user’s audio file, ‘*maayo’* and saving it to be analyzed by the voice comparison algorithm which is a mixture of the five algorithms: Silence Removal, Framing, Windowing, Linear Predictive Coding and Dynamic Time Warping.

**Pattern Matching**

**End**

**Processing**

**Preprocessing**

Silence

Removal, Noise Reduction

Framing and Windowing

**Start**

Capture and Save Audio

Decision

Making

Word Translation

Audio files

**Feature Extraction**

Linear Predictive

Coding

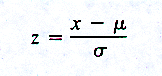
Dynamic Time Warping

**Database**

Figure 4.1 General Architecture

The speech signal in the audio file undergoes processing. Under the processing stage, the ‘*maayo’* signal is preprocessed first by taking out any existing silence portion and chopping the signal into frame and windows for easy processing. After preprocessing, the new signal’s important feature (user’s feature) is extracted using Linear Predictive Coding. The user’s feature is ready for the next stage: pattern matching. Inside pattern matching, the user’s feature is on wait, while going back to the processing stage. In the second passing of the processing stage, the audio files for ‘*maayo’* that are needed to be compared to the user’s feature are taken and will go through preprocessing and feature extraction. After that, the database’s feature of ‘*maayo’* proceeds to the last stage where the user’s feature of ‘*maayo’* is waiting. With Dynamic Time Warping, the two features will be compared and analyzed. Once a decision is being done, the algorithm displays the shortest distance or warping path for both ‘*maayo’* files.

## 4.1 Silence Removal and Noise Reduction

The audio file produces only one variable continuously which also means that the data (audio file) is univariate. A scale is used to measure the nearness and farness of the univariate data from the mean of an observation. This scale is often associated with “standard deviation” (Wicklin, 2012). The distance —nearness and farness — can be computed using the z-score:

Where x is the observation, µ is the mean, and σ is the standard deviation. In SR, the bigger the value of the z-score, the smaller possibility of the variable to be classified as voiced (Saha, et al., 2005).

One method to remove noise is by smoothing operation. A [low pass filter](http://en.wikipedia.org/wiki/Low_pass_filter) works by blocking the higher frequency (noise) and allowing the low frequency signal to pass (Eaton, S., 2009). Low-pass filters provide a smoother form of a signal, removing the short-term fluctuations, and leaving the longer-term trend.

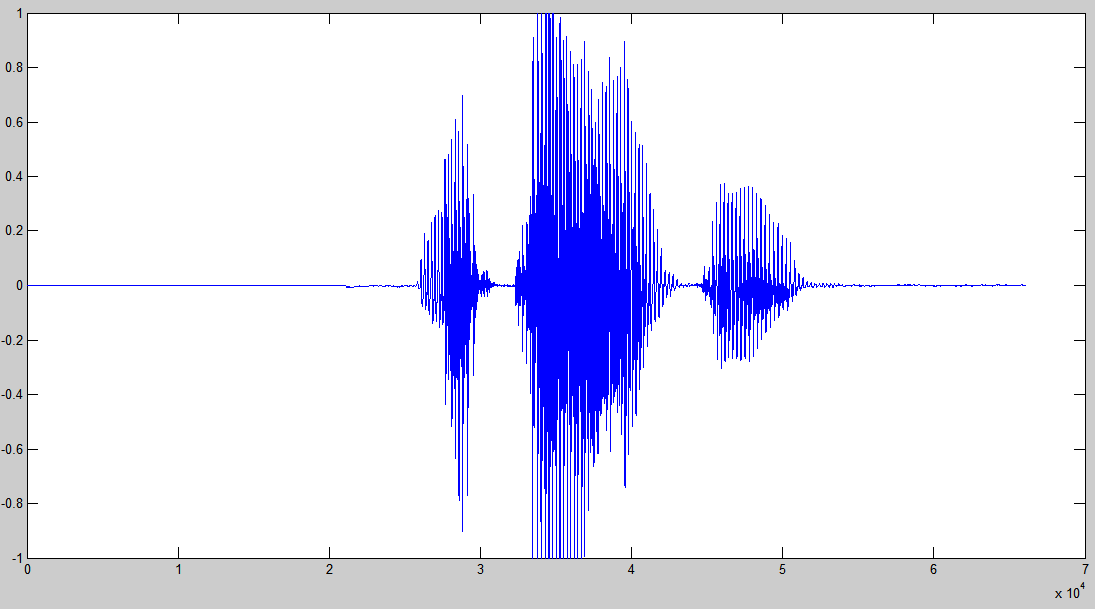


Figure 4.1.1 Raw file of ‘maayo’

Figure 4.1.1 is the untouched audio file of *maayo*, saved after being recorded. It has a long streak of silence at the first part of the signal. The silence is useless information and it has to be omitted or taken out as to immediately compare the files to where the actual voice starts.

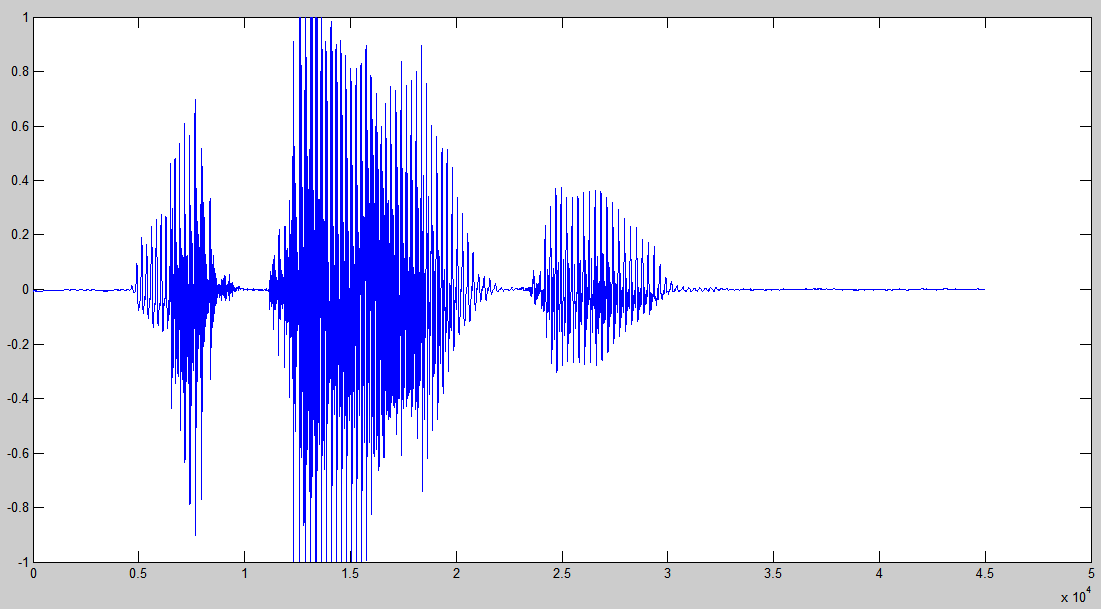


Figure 4.1.2 ‘maayo’ file after silence removal

After tampering with the *‘maayo’* audio file, Figure 4.1.2 shows how the silence is erased, to make the actual voice start at time 0.

The segmentation of a waveform of an audio signal is not exact. At 10 or 20 millissecond or less, that segment can be a product of silence, a voiced speech, unvoiced speech or noise. Small errors are welcomed as unvoiced part is studied to have low energy content and is thus categorized similarly with silence (Saha, et al., 2005). Figure 4.1.1 contains a long stretch of silence from time zero to 2x10-4 on the x-axis. The silence is denoted by the constant point of zero value. SR then creates only two classifications: voiced and unvoiced/silence. Voiced classification is a result of tensed vocal chords, producing vibration periodically. Unvoiced classification is when the vocal cords do not vibrate as silence is the absence of sound. After going through the selection of voiced or unvoiced, the audio file evidently gets smaller (refer to Figure 4.1.2 for the difference).

## 4.2 Frame Blocking and Hamming Window

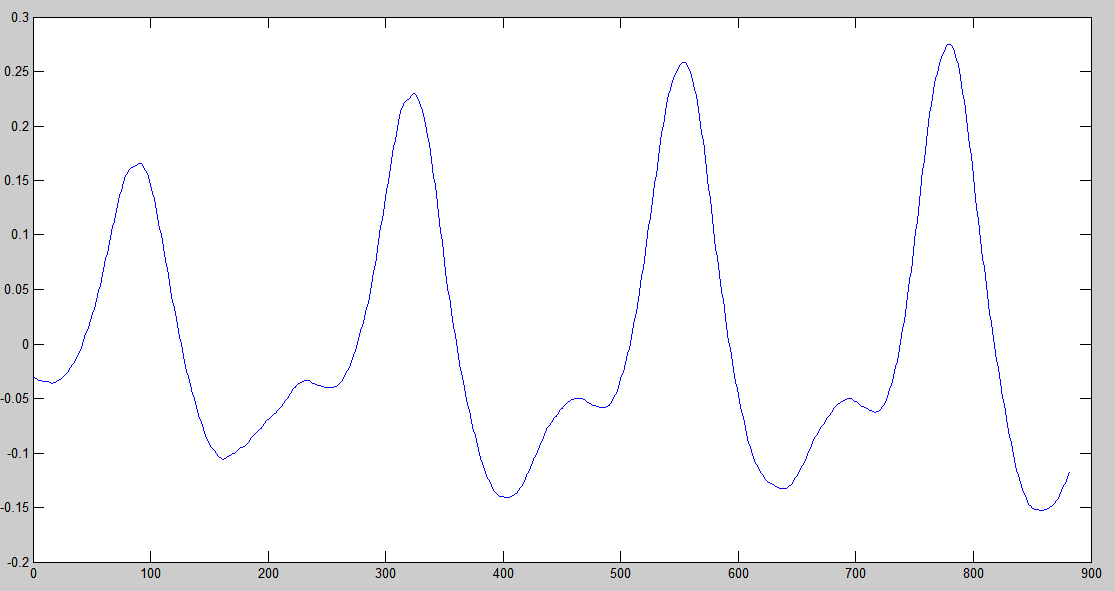
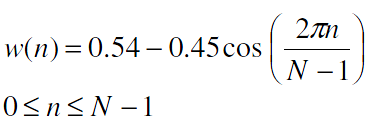


Figure 4.2.1 One Frame Block of ‘maayo’ file

Figure 4.2.1 is a closer view of the audio file, which is one frame.

Speech needs to be a stationary signal for easy analysis. Chopping the signal to frame blocks becomes short enough to be considered as stationary, like what is seen in Figure 4.2.1. The properties of the speech in a framed signal preferably must not change so much and also long enough to give substantial information about the frame (Çelik, 2008). Speech is typically analyzed between ten milliseconds and 30 milliseconds. For this algorithm, 20 milliseconds is the chosen time interval.

To minimize discontinuity and therefore preventing leakage or breakage on the signal at the beginning and end of each frame, every frame is multiplied by windowing function (Mosa, et al., 2009). The commonly used function is the Hamming window (Navarro, 2007):



Where w(n) is the audio signal and N represents the sample per frame. Figure 4.2.2 shows the output after windowing.

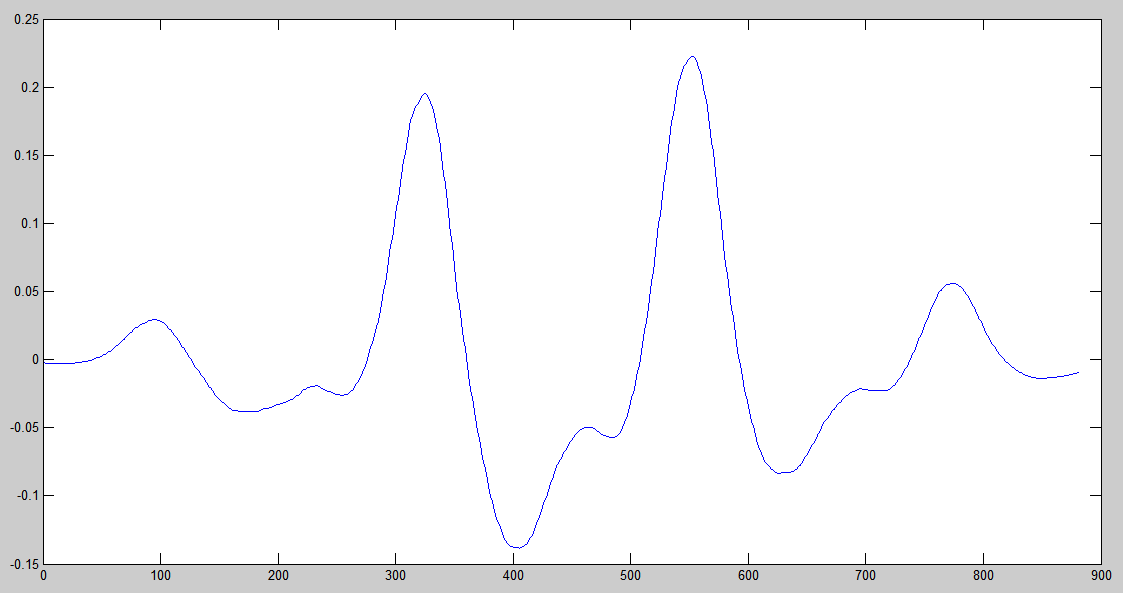
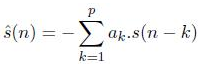


Figure 4.2.2 ‘maayo’ signal after windowing

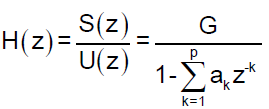
Figure 4.2.2 is the output of a framed signal multiplied by the Hamming window.

## 4.3 Feature Extraction – Linear Predictive Coding

The Linear Predictive Coding (LPC) captures the correlation of a variable— a sample of an audio file— in a form of prediction coefficients aq (Smith, 2011). The order of the coefficients p ranges from eight to 13. At each ten or 20 ms sample s(n) of an audio file, the number of coefficients produced is the order (Ambikairajah, 2010). The prediction sample ŝ(n) is taken.



These coefficients are used to separate speech signals into two parts: the transfer function (which contains the vocal quality) and the excitation (which contains the pitch and the sound). The input signal is an excitation of impulse train for voiced sound and random Gaussian noise for unvoiced sound. The switch rules out the signal that has been excited and scaled by a gain G. The signal enters into an all-pole IIR filter, which determines the characteristic of sound. The transfer function in the z-domain denoted as H(z) of the all pole AR(p) Infinite Impulse Response filter is given as



where s[n] is the windowed signal, u[n] is the excitation signal, G is the zero-frequency gain(Navarro, 2007).

## 4.4 Pattern Matching – Dynamic Time Warping

Dynamic Time Warping (DTW) is a technique that finds an optimal alignment between two time series in which one time series may be “warped” non-linearly by stretching or shrinking its time axis. This alignment can be used to find corresponding regions or to determine the similarity between the two time series (Salvador, et al., 2007). The coefficients of two signals are used as time series X and Y of lengths |X| and |Y|,

X = x1,x2, … x|X|

Y = y1,y2, … y|Y|

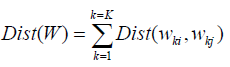
To construct a warp path W,

W = w1,w2, … wK

with K as length of the warp path and the kth element of the warp path is

wk = (i,j)

where i is an index of time series X, and j is an index of time series Y. The warp path starts at w1 = (1,1) until wk=(|X|, |Y|). An optimal warp path is a minimum distance warp path, where the distance (or cost) of a warp path W is



Dist(wki, wkj) is the distance between the two data point indexes (one from X and one from Y) in the kth element of the warp path. To find a minimum distance warp path, every cell in the cost matrix must be filled. If the solutions are already known for all slightly smaller portions of that time series that are a single data point away from lengths i and j, then the value at D(i, j) is the minimum distance for all these smaller time series, plus the distance between the points ii and jj. Since the warp path must either increase by one or stay the same along the i and j axes, the distances of the optimal warp paths one data point smaller than lengths i and j are contained in the matrix at D(i-1, j), D(i, j-1), and D(i-1, j-1). So the value of a cell in the cost matrix is

D(i, j) = Dist(i, j) +min[D(i −1, j), D(i, j −1), D(i −1, j −1)]

The warp path to D(i, j) must pass through one of those three cells, and since the minimum warp path distance is already known for them, all that is needed is to add the distance between the current pair of points, Dist(i, j), to the smallest value in those three cells. The cost matrix is filled one column at a time from the bottom up, from left to right. After the entire matrix is filled, a warp path must be found from D(1, 1) to D(|X|, |Y|). The warp path is calculated backwards, starting at D(|X|, |Y|). A search evaluates three nearby cells: to the left, below, and diagonally to the bottom-left. Whichever of these three cells has the smallest value is then added to the beginning of the warp path, and the search continues from that cell until D(1, 1) is reached.

## 4.5 Data Source

The idea was to use the most common words spoken in the English Language. The Oxford English Corpus (Oxford University Press, 2012) which currently contains two billion words, provided a list of 100 commonest English words being used around the world found on Table 4.5.1.

Table 4.5.1 100 Commonest English Words

|  |  |  |  |
| --- | --- | --- | --- |
| 1 the 2     be 3     to 4     of 5     and 6     a 7     in 8     that 9     have 10    I 11    it 12    for 13    not 14    on 15    with 16    he 17    as 18    you 19    do 20    at 21    this 22    but 23    his 24    by 25    from | 26    they 27    we 28    say 29    her 30    she 31    or 32    an 33    will 34    my 35    one 36    all 37    would 38    there 39    their 40    what 41    so 42    up 43    out 44    if 45    about 46    who 47    get 48    which 49    go 50    me | 51    when 52    make 53    can 54    like 55    time 56    no 57    just 58    him 59    know 60    take 61    people 62    into 63    year 64    your 65    good 66    some 67    could 68    them 69    see 70    other 71    than 72    then 73    now 74    look 75    only | 76    come 77    its 78    over 79    think 80    also 81    back 82    after 83    use 84    two 85    how 86    our 87    work 88    first 89    well 90    way 91    even 92    new 93    want 94    because 95    any 96    these 97    give 98    day 99    most 100   us |

The most frequently used words are short ones with the purpose to join other, longer words rather than determine the meaning of a sentence. These are known as 'function words'. The Oxford English Corpus explores the frequency of 'content words' from Table 4.5.1 and divided them by their categories shown on Table 4.5.2.

Table 4.5.2 List of Most Common Content Words

|  |  |  |
| --- | --- | --- |
| **Nouns** | **Verbs** | **Adjectives** |
| 1       time 2       person 3       year 4       way 5       day 6       thing 7       man 8       world 9       life 10      hand 11      part 12      child 13      eye 14      woman 15      place 16      work 17      week 18      case 19      point 20      government 21      company 22      number 23      group 24      problem 25      fact | 1       be 2       have 3       do 4       say 5       get 6       make 7       go 8       know 9       take 10      see 11      come 12      think 13      look 14      want 15      give 16      use 17      find 18      tell 19      ask 20      work 21      seem 22      feel 23      try 24      leave 25      call | 1       good 2       new 3       first 4       last 5       long 6       great 7       little 8       own 9       other 10      old 11      right 12      big 13      high 14      different 15      small 16      large 17      next 18      early 19      young 20      important 21      few 22      public 23      bad 24      same 25      able |

The nouns found in the list form parts of common phrases: some of the frequency of time, for example, comes from its use in adverbial phrases like on time, last time, next time and many more. The commonest verbs express basic concepts. In the list, most of these verbs are one-syllabic. Most of the top adjectives are also one-syllabic and may be synonymic to the other words for the reason that a large choice is available for expression.

## 4.6 Database Model

A simple database of Cebuano words are categorized under their respective form and their corresponding English translation, found on Table 4.6.1.

Table 4.6.1 Singular Row of the Database

|  |  |  |  |
| --- | --- | --- | --- |
| ID | Cebuano | English | Form |

Each Cebuano word has ten audio files placed in folder with the name of the Cebuano word. These audio files were recorded through a java code with a sampling rate of 44100 kHz and a sample size of 16 bits in the F0120 Physics Lab Control room of the University of San Carlos. The microphone has a frequency response of 20 Hz to 20 kHz with a rated power of 100mW and sensitivity of 103 dB S.P.L at 7 kHz.

The checking of the Cebuano pronunciation of the words are checked by Miss Cindy Velasquez, a Humanities and Literature professor of the University of San Carlos and a member of Bathalan-ong Halad sa Dagan and Women in Literary Arts.

## 4.7 Definition of Terms

For clearer understanding of the terms used in this study, below are their meanings:

**Audio Signal** – It is a representation of [sound](http://en.wikipedia.org/wiki/Sound), typically as an electrical [voltage](http://en.wikipedia.org/wiki/Voltage).

**Frequency** – It is the rate at which a vibration occurs that constitutes a wave, in a material (as in sound waves).

**Linear Prediction** – It is a mathematical operation where future values of a [discrete-time](http://en.wikipedia.org/wiki/Discrete_time) [signal](http://en.wikipedia.org/wiki/Signal_processing) are estimated as a [linear function](http://en.wikipedia.org/wiki/Linear_transformation) of previous samples.

**Noise Reduction –** It is the process of removing noise from a signal.

**Pattern Matching** – It refers to the act of checking a perceived sequence of tokens for the presence of the constituents of some [pattern](https://en.wikipedia.org/wiki/Pattern).

**Pronunciation Checking –** It is the process of listening to a certain word, practicing the word, and testing the correct pronunciation.

**Time Series** – It is a series of values of a quantity obtained at successive times, often with equal intervals between them.

**Use Case Diagram** – is a graphical representation of the user's interaction with the system.

**Waveform** - It means the shape and form of a [signal](http://en.wikipedia.org/wiki/Signal_(information_theory)) such as a [wave](http://en.wikipedia.org/wiki/Wave) moving in a physical medium or an abstract representation.

# CHAPTER 5

# RESULTS AND ANALYSIS

## 5.1 Softwares Used

The XAMPP is used for the database system. XAMPP is a free package available for download and use for various development tasks. It provides support for administering and maintaining MySQL databases. Considering the possibility of a future research, relating to this research and/or having to reuse the database, the main benefits of XAMPP such as its easy configuration, easy to read documentation, and free distribution in a hope that it will be useful, will be utilized.

The software used is the NetBeans IDE since all the algorithms in this research are implemented in the Java language. Having a bewildering amount of languages to choose from, Java is mainly at an advantage because of its availability on many platforms and simplicity of syntax. For future projects that will extend the scope of this research, the Java knowledge easily transfers to a multitude of other available languages, together with Java’s syntax enhancements.

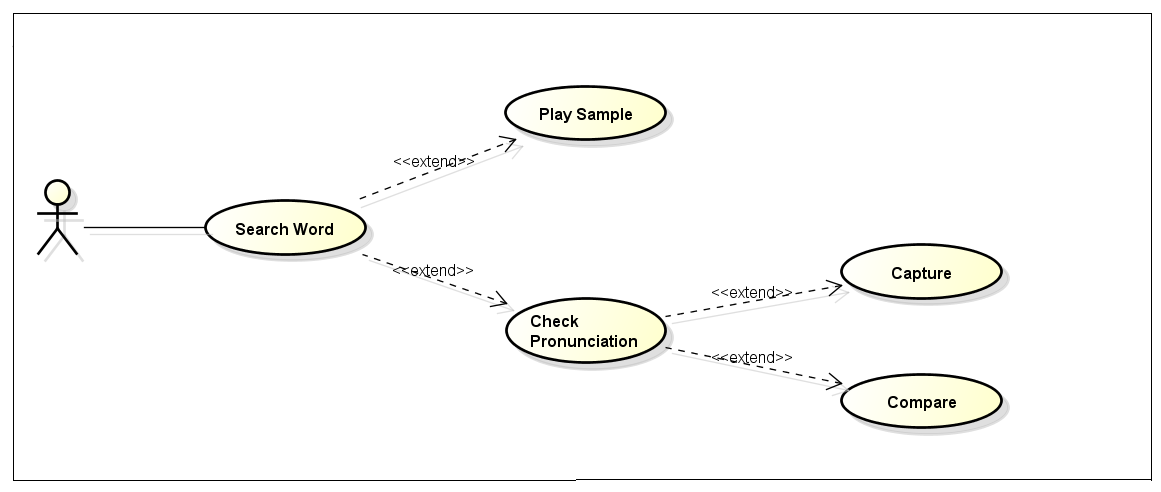


Figure 5.1 Use Case Diagram

Figure 5.1 is the diagram of the user and the functions he can do. The first thing he can do is to search for an English word available in the database. He can then decide whether to play the sample or compare it, as the use case *search word* extends *play sample* and *check pronunciation*. Inside *check pronunciation* two conditionings are given: *capture* his voice or *compare* his captured voice to the audio files in the database.

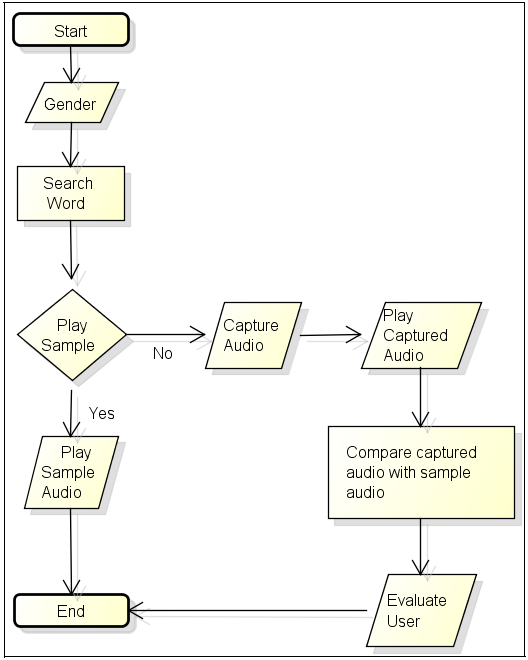


Figure 5.2 Control Flow Diagram

Figure 5.2 shows the actual flow of the system; where it starts and how it ends.

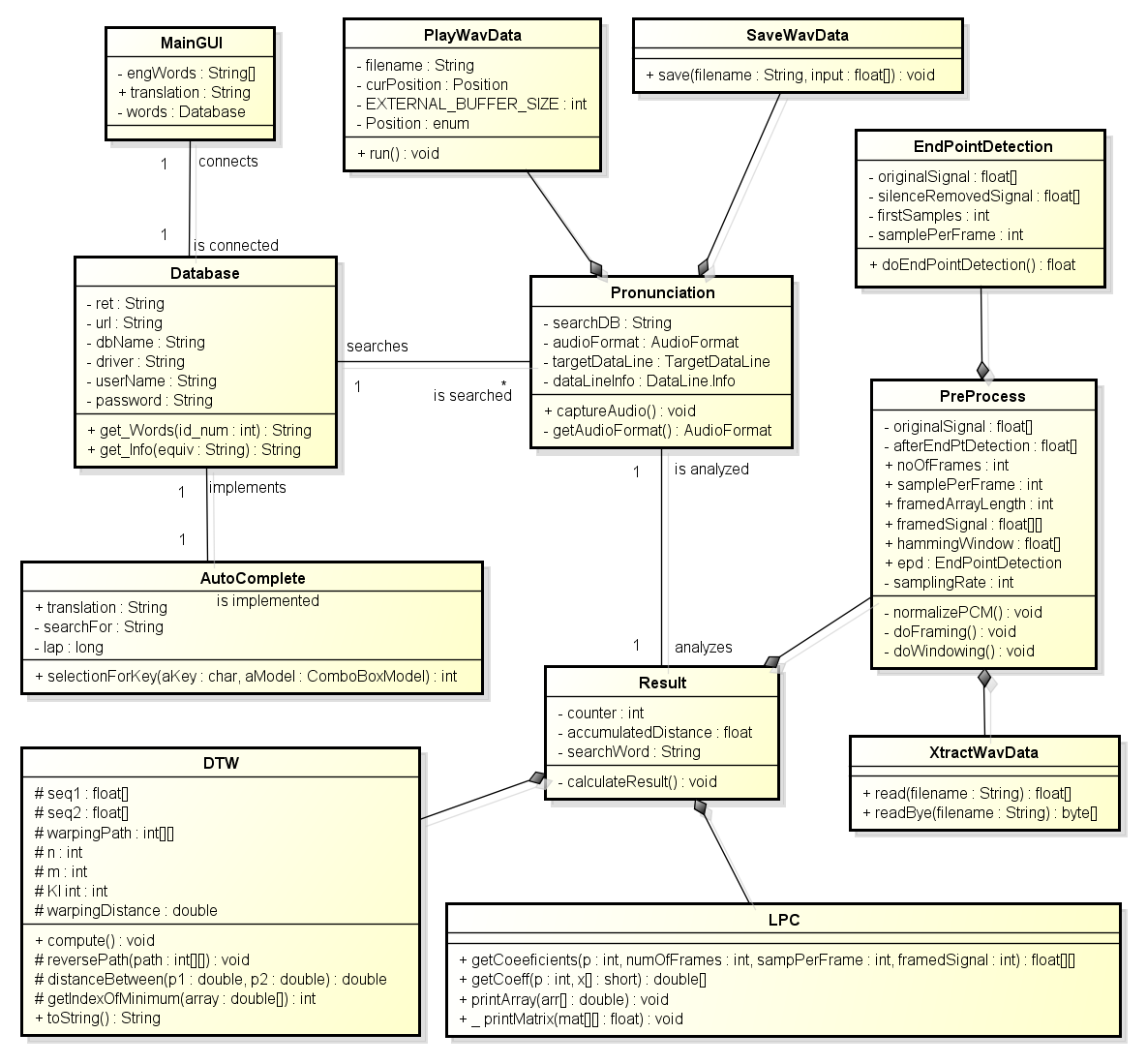
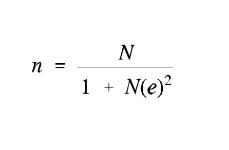


Figure 5.3 Class Diagram

Figure 5.3 is the diagram of the classes we used, the variables and methods that are contained in the system.

## 5.2 Determining Sample Size

Based on the evidence of the Oxford English Corpus (2012), instead of talking about words, a lemma being the base form of a word is more useful in determining the commonest English words found in writing around the world. From the 100 commonest words listed, this study uses the following formula to determine the total number of words needed:

[](http://edis.ifas.ufl.edu/LyraEDISServlet?command=getImageDetail&image_soid=IMAGE%20PD:PD006E3A&document_soid=PD006&document_version=56091)

Where n is the sample size and N is the population size, and e is the level of precision. There are 100 words listed in the Oxford Dictionaries, Thus, N=100. With a sampling error of ±10% where risk level is 95%, n can be computed by:

n = 100 = 50 words

1 + 100 ( 0.10 ) 2

## 5.3 Test Results

Researchers who are interested in creating interval scales (scales in which the respondents perceive equal-sized gradations between the points on the scale) must be careful to choose category descriptors that are truly equal-interval. Ideally, a rating scale should consist of enough points to extract the necessary information. When a researcher is interested in averages across people or the combination of several rating scale, two- or three- point scales (Friedman, et al, 1999). Jacoby and Mattel (1972) claims that scales consisting of three points are sufficient to meet the criteria of predictive validity. This study uses a three-point interval scale as the type of rati ng scale.

To provide proportionately more information and to determine the effectiveness of the algorithm in recognizing the words that are pronounced correctly, the following tables present the results of the test done with two adults of the both gender, and children of both gender. This study uses the following formula for the number of words for each tester:

Z2 \* p \* (1-p)

n0 =

C2

Where n0 is the sample size, z is the z-score, p is the percentage and C is the confidence interval. The standard z-score for a 95% confidence level is 1.96 (taken from a statistical z-score table), percentage is 0.5 (expressed in decimal), and C, the confidence interval is 21%. n0 can be computed by:

1.962 \* 0.5 \* (1-0.5)

n0 = = 21.77778

0.212

With the population being small, the sample size can be reduced slightly. The n0 can be adjusted using the following formula:

n0

n =

1+ (n0 – 1)

N

Where n is the new sample size, and N is the population size. With N = 50, the new sample size can be adjusted to:

21.77778

n = = 15 words

1+ (21.77778– 1)

50

Table 5.3.1 Test Results: Female Adult

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Input** | **Expected Result** | **Actual Result** | **Actual Value** | **Status** |
| aduna | Average | Good | 0.4268251 | fairly acceptable |
|  | Good | Good | 0.4923096 | completely acceptable |
|  | Excellent | Excellent | 0.30711565 | completely acceptable |
|  | Poor | Poor | 1.5552042 | completely acceptable |
| buhat | Average | Average | 0.75026923 | completely acceptable |
|  | Good | Average | 0.82695895 | fairly acceptable |
|  | Excellent | Good | 0.546496 | fairly acceptable |
|  | Poor | Average | 0.99358577 | unacceptable |
| sulti | Average | Average | 1.0912894 | completely acceptable |
|  | Good | Average | 1.0337569 | fairly acceptable |
|  | Excellent | Excellent | 0.248712 | completely acceptable |
|  | Poor | Poor | 1.2297895 | completely acceptable |
| panag-iya | Average | Average | 1.0090563 | completely acceptable |
|  | Good | Average | 0.9216 | fairly acceptable |
|  | Excellent | Average | 0.9921 | unacceptable |
|  | Poor | Poor | 1.43824 | completely acceptable |
| kuha | Average | Average | 0.721817 | completely acceptable |
|  | Good | Average | 0.6586056 | fairly acceptable |
|  | Excellent | Good | 0.40664676 | fairly acceptable |
|  | Poor | Poor | 1.1326 | completely acceptable |
| buhat | Average | Good | 0.54657984 | fairly acceptable |
|  | Good | Good | 0.3534552 | completely acceptable |
|  | Excellent | Excellent | 0.24568 | completely acceptable |
|  | Poor | Average | 0.99453 | unacceptable |
| adto | Average | Average | 0.9070874 | completely acceptable |
|  | Good | Average | 0.91048515 | fairly acceptable |
|  | Excellent | Good | 0.587295 | fairly acceptable |
|  | Poor | Poor | 1.46534 | completely acceptable |
| kahibawo | Average | Average | 0.67455 | completely acceptable |
|  | Good | Good | 0.41896 | completely acceptable |
|  | Excellent | Excellent | 0.31726 | completely acceptable |
|  | Poor | Average | 1.0693 | unacceptable |
| kita | Average | Average | 0.713597 | completely acceptable |
|  | Good | Good | 0.44174253 | completely acceptable |
|  | Excellent | Excellent | 0.15956813 | completely acceptable |
|  | Poor | Poor | 1.67595 | completely acceptable |
| duol | Average | Average | 1.064134 | completely acceptable |
|  | Good | Average | 0.8231844 | fairly acceptable |
|  | Excellent | Excellent | 0.14739785 | completely acceptable |
|  | Poor | Average | 1.828457 | unacceptable |
| hunahuna | Average | Poor | 1.2081365 | unacceptable |
|  | Good | Good | 0.44985 | completely acceptable |
|  | Excellent | Excellent | 0.217511 | completely acceptable |
|  | Poor | Poor | 2.125109 | completely acceptable |
| hatag | Average | Average | 0.99684 | completely acceptable |
|  | Good | Good | 0.545984 | completely acceptable |
|  | Excellent | Excellent | 0.3146341 | completely acceptable |
|  | Poor | Average | 1.070943 | unacceptable |
| gamit | Average | Average | 0.623498 | completely acceptable |
|  | Good | Good | 0.34689734 | completely acceptable |
|  | Excellent | Excellent | 0.2227182 | completely acceptable |
|  | Poor | Poor | 1.613849 | completely acceptable |
| pagpangita | Average | Average | 0.90913 | completely acceptable |
|  | Good | Good | 0.517478 | completely acceptable |
|  | Excellent | Excellent | 0.4619845 | completely acceptable |
|  | Poor | Poor | 1.423112 | completely acceptable |
| oras | Average | Good | 0.5153398 | fairly acceptable |
|  | Good | Good | 0.5314982 | completely acceptable |
|  | Excellent | Excellent | 0.10995 | completely acceptable |
|  | Poor | Poor | 1.2334122 | completely acceptable |

Table 5.3.1 contains four columns (not included the ‘word column’), where: Expected Result is the rating predicted, Actual Result is the rating returned by the system, Actual Value is the test value, and Status is the qualitative attribute. The test is to pronounce the words correctly and estimate-by-ear near to the database audio file. These are the calculated distance cost for each warping of the path, until it reaches D(i,j) wherein i is the last column and j is the first row. Refer to the section 4.4 under 4.0 The Algorithms and their Models.

Table 5.3.2 Test Results: Male Adult

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Input** | **Expected Result** | **Actual Result** | **Actual Value** | **Status** |
| aduna | Average | Average | 0.83019454 | completely acceptable |
|  | Good | Good | 0.5671323 | completely acceptable |
|  | Excellent | Excellent | 0.3185921 | completely acceptable |
|  | Poor | Poor | 1.2113886 | completely acceptable |
| buhat | Average | Average | 1.7289205 | unacceptable |
|  | Good | Average | 0.8056584 | fairly acceptable |
|  | Excellent | Good | 0.59534693 | fairly acceptable |
|  | Poor | Poor | 1.6257351 | completely acceptable |
| sulti | Average | Poor | 1.106731 | unacceptable |
|  | Good | Good | 0.3513447 | completely acceptable |
|  | Excellent | Excellent | 0.24787 | completely acceptable |
|  | Poor | Poor | 1.81769 | completely acceptable |
| panag-iya | Average | Average | 0.993613 | completely acceptable |
|  | Good | Excellent | 0.313434 | fairly acceptable |
|  | Excellent | Good | 0.523565 | fairly acceptable |
|  | Poor | Average | 0.923826 | unacceptable |
| kuha | Average | Average | 0.994213 | completely acceptable |
|  | Good | Good | 0.613295 | completely acceptable |
|  | Excellent | Good | 0.464613 | fairly acceptable |
|  | Poor | Poor | 1.223398 | completely acceptable |
| buhat | Average | Average | 0.7896512 | completely acceptable |
|  | Good | Average | 0.6867351 | fairly acceptable |
|  | Excellent | Good | 0.609256 | fairly acceptable |
|  | Poor | Poor | 1.4377118 | completely acceptable |
| adto | Average | Average | 0.8936803 | completely acceptable |
|  | Good | Average | 0.82574871 | fairly acceptable |
|  | Excellent | Excellent | 0.295887 | completely acceptable |
|  | Poor | Poor | 1.324565 | completely acceptable |
| kahibawo | Average | Average | 0.7546949 | completely acceptable |
|  | Good | Good | 0.41952 | completely acceptable |
|  | Excellent | Excellent | 0.312508 | completely acceptable |
|  | Poor | Poor | 1.92687 | completely acceptable |
| kita | Average | Average | 0.7211236 | completely acceptable |
|  | Good | Excellent | 0.16363468 | fairly acceptable |
|  | Excellent | Excellent | 0.1591996 | fairly acceptable |
|  | Poor | Poor | 1.7134 | completely acceptable |
| duol | Average | Average | 0.9707128 | completely acceptable |
|  | Good | Good | 0.2971578 | fairly acceptable |
|  | Excellent | Excellent | 0.2404575 | completely acceptable |
|  | Poor | Poor | 1.54645 | completely acceptable |
| hunahuna | Average | Poor | 1.19897 | unacceptable |
|  | Good | Average | 0.67234 | completely acceptable |
|  | Excellent | Excellent | 0.313457 | completely acceptable |
|  | Poor | Average | 1.04434 | unacceptable |
| hatag | Average | Average | 0.825976 | completely acceptable |
|  | Good | Good | 0.59868 | completely acceptable |
|  | Excellent | Good | 0.48958 | fairly acceptable |
|  | Poor | Poor | 1.209997 | completely acceptable |
| gamit | Average | Good | 0.56939 | fairly acceptable |
|  | Good | Good | 0.5173413 | completely acceptable |
|  | Excellent | Good | 0.423431 | fairly acceptable |
|  | Poor | Poor | 1.498705 | completely acceptable |
| pagpangita | Average | Average | 0.921566 | completely acceptable |
|  | Good | Average | 0.63316 | fairly acceptable |
|  | Excellent | Excellent | 0.309098 | completely acceptable |
|  | Poor | Poor | 1.8392 | completely acceptable |
| oras | Average | Good | 0.52096 | fairly acceptable |
|  | Good | Good | 0.56187 | completely acceptable |
|  | Excellent | Excellent | 0.173492 | completely acceptable |
|  | Poor | Poor | 2.197783 | completely acceptable |

Table 5.3.2 contains the test results for an adult male with four columns (not included the ‘word column’), where: Expected Result is the rating predicted, Actual Result is the rating returned by the system, Actual Value is the test value, and Status is the qualitative attribute.

Table 5.3.3 Test Results: Female Child

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Input** | **Expected Result** | **Actual Result** | **Actual Value** | **Status** |
| aduna | Average | Average | 0.984759 | completely acceptable |
|  | Good | Good | 0.59877 | completely acceptable |
|  | Excellent | Good | 0.39087 | fairly acceptable |
|  | Poor | Poor | 2.37483 | completely acceptable |
| buhat | Average | Average | 0.97685 | completely acceptable |
|  | Good | Average | 0.84876 | fairly acceptable |
|  | Excellent | Average | 0.87655 | fairly acceptable |
|  | Poor | Poor | 1.87688 | completely acceptable |
| sulti | Average | Good | 0.57534 | fairly acceptable |
|  | Good | Average | 0.8765 | fairly acceptable |
|  | Excellent | Excellent | 0.31886 | completely acceptable |
|  | Poor | Poor | 2.37459 | completely acceptable |
| panag-iya | Average | Average | 1.08365 | completely acceptable |
|  | Good | Good | 0.61975 | completely acceptable |
|  | Excellent | Average | 0.78759 | fairly acceptable |
|  | Poor | Poor | 1.985709 | completely acceptable |
| kuha | Average | Average | 0.893426 | completely acceptable |
|  | Good | Excellent | 0.26532 | fairly acceptable |
|  | Excellent | Excellent | 0.29836 | completely acceptable |
|  | Poor | Poor | 1.76398 | completely acceptable |
| buhat | Average | Good | 0.77956 | fairly acceptable |
|  | Good | Good | 0.610979 | completely acceptable |
|  | Excellent | Average | 0.89457 | fairly acceptable |
|  | Poor | Poor | 1.9538 | completely acceptable |
| adto | Average | Poor | 1.19562 | unacceptable |
|  | Good | Excellent | 0.19392 | fairly acceptable |
|  | Excellent | Good | 0.45187 | fairly acceptable |
|  | Poor | Poor | 1.87103 | completely acceptable |
| kahibawo | Average | Good | 0.58451 | fairly acceptable |
|  | Good | Good | 0.59835 | completely acceptable |
|  | Excellent | Good | 0.49068 | fairly acceptable |
|  | Poor | Average | 1.109483 | fairly acceptable |
| kita | Average | Average | 0.98376 | completely acceptable |
|  | Good | Average | 0.627356 | fairly acceptable |
|  | Excellent | Excellent | 0.305619 | completely acceptable |
|  | Poor | Poor | 2.104538 | completely acceptable |
| duol | Average | Poor | 1.96794 | fairly acceptable |
|  | Good | Average | 0.6714356 | fairly acceptable |
|  | Excellent | Excellent | 0.24972 | fairly acceptable |
|  | Poor | Average | 1.039476 | fairly acceptable |
| hunahuna | Average | Poor | 1.561398 | unacceptable |
|  | Good | Good | 0.45297 | completely acceptable |
|  | Excellent | Good | 0.398652 | fairly acceptable |
|  | Poor | Poor | 1.56034 | completely acceptable |
| hatag | Average | Average | 0.86547 | completely acceptable |
|  | Good | Good | 0.530976 | completely acceptable |
|  | Excellent | Average | 0.86278 | fairly acceptable |
|  | Poor | Poor | 1.846832 | completely acceptable |
| gamit | Average | Average | 0.986352 | completely acceptable |
|  | Good | Average | 0.95794 | completely acceptable |
|  | Excellent | Excellent | 0.198453 | completely acceptable |
|  | Poor | Poor | 1.907893 | completely acceptable |
| pagpangita | Average | Poor | 1.36894 | unacceptable |
|  | Good | Average | 0.679538 | fairly acceptable |
|  | Excellent | Good | 0.479573 | fairly acceptable |
|  | Poor | Poor | 1.49867 | completely acceptable |
| oras | Average | Poor | 1.4986 | unacceptable |
|  | Good | Good | 0.40968 | completely acceptable |
|  | Excellent | Excellent | 0.30761 | completely acceptable |
|  | Poor | Poor | 2.94814 | completely acceptable |

Table 5.3.3 contains the test results for a female child with four columns (not included the ‘word column’), where: Expected Result is the rating predicted, Actual Result is the rating returned by the system, Actual Value is the test value, and Status is the qualitative attribute.

**Table 5.3.4 Test Results: Male Child**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Input** | **Expected Result** | **Actual Result** | **Actual Value** | **Status** |
| aduna | Average | Good | 0.4024646 | fairly acceptable |
|  | Good | Average | 0.7507524 | fairly acceptable |
|  | Excellent | Excellent | 0.1349755 | completely acceptable |
|  | Poor | Poor | 1.954739 | completely acceptable |
| buhat | Average | Poor | 1.9784639 | unacceptable |
|  | Good | Good | 0.46383 | completely acceptable |
|  | Excellent | Good | 0.4756593 | fairly acceptable |
|  | Poor | Poor | 2.846592 | completely acceptable |
| sulti | Average | Poor | 1.13846 | unacceptable |
|  | Good | Good | 0.464037 | completely acceptable |
|  | Excellent | Good | 0.485781 | fairly acceptable |
|  | Poor | Poor | 1.104855 | completely acceptable |
| panag-iya | Average | Poor | 1.1048574 | unacceptable |
|  | Good | Good | 0.495721 | completely acceptable |
|  | Excellent | Average | 0.7659 | fairly acceptable |
|  | Poor | Average | 1.9548 | unacceptable |
| kuha | Average | Average | 0.6482 | completely acceptable |
|  | Good | Good | 0.6136 | completely acceptable |
|  | Excellent | Excellent | 0.2948531 | completely acceptable |
|  | Poor | Average | 1.04735 | unacceptable |
| buhat | Average | Average | 0.98163 | completely acceptable |
|  | Good | Average | 0.69581 | fairly acceptable |
|  | Excellent | Excellent | 0.1349 | completely acceptable |
|  | Poor | Poor | 1.20561 | completely acceptable |
| adto | Average | Good | 0.4345702 | fairly acceptable |
|  | Good | Average | 0.8319978 | fairly acceptable |
|  | Excellent | Excellent | 0.231929 | completely acceptable |
|  | Poor | Poor | 1.47943 | completely acceptable |
| kahibawo | Average | Poor | 1.1438317 | unacceptable |
|  | Good | Good | 0.61235 | completely acceptable |
|  | Excellent | Good | 0.4806444 | fairly acceptable |
|  | Poor | Poor | 1.266553 | completely acceptable |
| kita | Average | Average | 0.8694985 | completely acceptable |
|  | Good | Excellent | 0.31672 | fairly acceptable |
|  | Excellent | Average | 0.798025 | fairly acceptable |
|  | Poor | Poor | 1.116469 | completely acceptable |
| duol | Average | Good | 0.61939 | fairly acceptable |
|  | Good | Good | 0.561192 | completely acceptable |
|  | Excellent | Excellent | 0.16368 | completely acceptable |
|  | Poor | Poor | 1.051668 | completely acceptable |
| hunahuna | Average | Average | 0.7693976 | completely acceptable |
|  | Good | Excellent | 0.140607 | fairly acceptable |
|  | Excellent | Good | 0.453417 | fairly acceptable |
|  | Poor | Average | 1.01831 | unacceptable |
| hatag | Average | Average | 0.827438 | completely acceptable |
|  | Good | Average | 0.866943 | fairly acceptable |
|  | Excellent | Good | 0.332168 | fairly acceptable |
|  | Poor | Poor | 1.889 | completely acceptable |
| gamit | Average | Average | 0.85579 | completely acceptable |
|  | Good | Good | 0.395158 | completely acceptable |
|  | Excellent | Excellent | 0.172672 | completely acceptable |
|  | Poor | Average | 0.92196 | fairly acceptable |
| pagpangita | Average | Average | 0.824193 | completely acceptable |
|  | Good | Average | 0.6892591 | fairly acceptable |
|  | Excellent | Good | 0.4613 | fairly acceptable |
|  | Poor | Poor | 1.2957328 | completely acceptable |
| oras | Average | Average | 0.9148342 | completely acceptable |
|  | Good | Good | 0.57868 | completely acceptable |
|  | Excellent | Good | 0.615023 | fairly acceptable |
|  | Poor | Poor | 1.7664219 | completely acceptable |

Table 5.3.4 contains the test results for a male child with four columns (not included the ‘word column’), where: Expected Result is the rating predicted, Actual Result is the rating returned by the system, Actual Value is the test value, and Status is the qualitative attribute.

**Table 5.3.5 Sample Mean Scores**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **sample mean score** | | | |
| **word** | **Average** | **Good** | **Excellent** | **Poor** |
| aduna | 0.66106081 | 0.602241075 | 0.28788831 | 1.77404045 |
| buhat | 1.358625908 | 0.736301838 | 0.62351306 | 1.835698218 |
| sulti | 0.9779551 | 0.68140965 | 0.32530575 | 1.631731125 |
| panag-iya | 1.047794175 | 0.58762625 | 0.76728875 | 1.57564375 |
| kuha | 0.814414 | 0.53770515 | 0.36611822 | 1.291832 |
| buhat | 0.77435526 | 0.586744825 | 0.4711015 | 1.39791295 |
| adto | 0.857739475 | 0.690537915 | 0.39174525 | 1.53509125 |
| kahibawo | 0.78939665 | 0.512295 | 0.4002731 | 1.3430515 |
| kita | 0.821994775 | 0.387363303 | 0.35560293 | 1.65258925 |
| duol | 1.1555442 | 0.58824245 | 0.20031384 | 1.36651275 |
| hunahuna | 1.184475525 | 0.42894175 | 0.34575925 | 1.46202475 |
| hatag | 0.878931 | 0.63564575 | 0.49979053 | 1.504193 |
| gamit | 0.7587575 | 0.55433416 | 0.25431855 | 1.48560175 |
| pagpangita | 1.00595725 | 0.629858775 | 0.42798888 | 1.5141787 |
| oras | 0.8624335 | 0.52043205 | 0.30151875 | 2.036439275 |

The data gathered on Table 5.3.5 of the second column (Average), third column (Good), fourth column (Excellent) and last column (Poor) are the calculated sample mean for the equivalent rating of the corresponding word. These data are the basis for the decision factor and evaluation of the algorithm.

## 5.4 Evaluation

In a study conducted by Eamonn Keogh and Michael Pazzani (2001), a classic Dynamic Time Warping’s (DTW) objective, even in scattered values, are normalized to have a mean of zero and a standard deviation of 1. It suggests that the acceptable values must be at most 1. Anything above it is considered not acceptable. The more warping discovered the larger the value of the mean.

The extracted test values were categorized according to acceptance rate. Actual Results that matched with the Expected Results fall under the completely acceptable category. Test values whose rating didn’t matched the Expected Result but still obtained the acceptable value, are considered to be fairly acceptable. Under unacceptable category are values whose Expected Result is Poor but actually passed, or vice versa.

**Figure 5.3.1 Results for Female Adult**

For the Female Adult, the acceptable test values have been found out to be 88%. Out of the 15 words, each with three-point interval testing and one fault testing, 41 test values have reached the objective (0.68333333%) with a completely acceptable scoring.

**Figure 5.3.2 Results for Male Adult**

As reflected in Figure 5.3.2, the overall acceptable rate for Male Adult was 92%. The number of test values that fell to the completely acceptable category is 39 out of 60.

**Figure 5.3.3 Results for Female Child**

In Figure 5.3.3, the overall acceptable rate for the Female Child was 93% but the number of values that are completely acceptable have reduced to more than 10%. Out of 60 tests, only 33 were completely acceptable.

**Figure 5.3.4 Results for Male Child**

Figure 5.3.4 described that the overall acceptable rate for the Male Child got 88%, and the completely acceptable rate also reduced. Only 33 out of 60 test values made it to the completely acceptable category.

**Figure 5.4 Overall Test Results Categorized to Three Acceptance Rate**

As observed in Table 5.3.4, the unacceptable actual value of the word ‘sulti’ which is 1.13846, is 0.1 more than the accepted scoring for ‘Average’ rating. Unacceptable scoring like ‘hunahuna’ in Table 5.3.2 with 1.04434 as the actual value is also 0.06 less than the accepted scoring for ‘Poor’ rating. The algorithm works completely accurate by 61% and 30% slightly accurate as seen in Figure 5.4.

**Figure 5.5 Overall Test Results by Three-Point Interval Scale**

Furthermore, as reflected in Figure 5.5, 68% of the mean scores of each word have reached the acceptable value.

**Figure 5.5 Overall Test Results**

The concept of classic Dynamic Time Warping means that a shorter warping path leads nearer to the acceptable area. Out of the 240 tests, 217 test values have reached the acceptable area, which is at most 1, disregarding the three interval scale used in this study. Overall, acceptable test values reached the objective (90%).

# CHAPTER 6

# CONCLUSION AND RECOMMENDATIONS

## 6.1 Conclusions

The algorithm works of about 68% in terms of accuracy categorized by a three-point interval rating scale. However, 90% is estimated to have obtained the correct pronunciation. It can be concluded that the results that the algorithm from this study give, denote accurate values that are significant enough in identifying the wrong pronunciation from the right one.

## 6.2 Recommendations

The researchers recommend further studies on LPC, by converting the LPC coefficients to cepstrum. More samples for the database on each word must be recorded to have higher accuracy. It is also best to minimize the sample rate, as to constrict the values of the signal. Dynamic Time Warping can further be studied, using fixed number of frames and not just the samples per frame. A faster pattern matching can be achieved by creating codebooks, rather than extracting the audio files at each comparison.

# APPENDIX A Curriculum Vitae

**CONTACT INFORMATION**

Name: Miraflor Codeniera Igot

Cell Phone: (+639) 225435198

Email: [miraigot@gmail.com](mailto:miraigot@gmail.com)

**EDUCATION**

Saint Alphonsus Catholic School, Lapu-Lapu City, Cebu

Highschool

March 2009

University of San Carlos, Cebu City, Cebu

Bachelor of Science in Computer Science

March 2013

**CONTACT INFORMATION**

Name: B Carla Obate Yap

Cell Phone: (+639)228612057

Email: [bcarlayap@ymail.com](mailto:bcarlayap@ymail.com)

**EDUCATION**

University of San Jose-Recoletos, Basak, Pardo, Cebu

Highschool

March 2009

University of San Carlos, Cebu City, Cebu

Bachelor of Science in Computer Science

March 2013

**CONTACT INFORMATION**

Name: Jacqueline Fat-Yara

Email: jgfat@yahoo.com

**EDUCATION**

University of the Philippines Visayas – Cebu High School, Salinas DriveLahug, Cebu

Highschool

March 1994

University of the Philippines Visayas, Salinas Drive Lahug, Cebu

Bachelor of Science in Math and Computer Science

March 1998

De La Salle University, Malate, Manila

Master of Science in Computer Science

**WORK EXPERIENCE**

**SPECIALIZATION**

# APPENDIX B Consultation Details

# References:

**Books:**

Heaton, G.B. (1988). Writing English Language Tests. London: Longman.

Proakis, J.G. and Manolakis, D.G. (2007). Introduction to Digital Signal Processing: Principles,

Algorithms, and Applications, 4th. ed. New York, Prentice Hall.

Räisänen, A. & Räisänen, N. (2007). The Alphabet and the Pronunciation.A handbook of

Cebuano. (pp. 7). Rochester, NY:Lulu.

Waibel, A. & Lee, K. (1990). Linear Predictive Coding. Readings in Speech Precognition. (pp. 60). United States, Morgan Kaufmann Publishers, Inc.

**Online/Internet:**

Adlawan, R. (2012). PB holds session in Cebuano. Sunstar Cebu. [online]. Available

[http://www.sunstar.com.ph/cebu/local-news/2012/02/28/pb-as-session-cebuano-208484](http://www.sunstar.com.ph/cebu/local-news/2012/02/28/pb-holds-session-cebuano-208484).

(March 16, 2012).

Aznar, M. (2012). China, HK arrivals down 40%. Sunstar Cebu. [online]. Available

<http://www.sunstar.com.ph/cebu/business/2012/02/24/china-hk-arrivals-down-40-207811>.

(March 16, 2012).

Çelik, Hüseyin. (2008). Speech Recognition. [online]. Available

http://www.scribd.com/doc/53236802/Speech-Recognition. (August 6, 2012).

Coleman J. (2011). Dynamic programming example (dynamic time warping). [online]. Available.

<http://www.phon.ox.ac.uk/jcoleman/old_SLP/Lecture_5/DTW_explanation.html>. (September 12, 2012).

Costanilla, R. & Montenegro, N. (2011). Probabilistic Speaker Identification to a Non-Continuous Text

Independent Verbal Communication. International Proceedings of Economics Development and Research. [online]. Availablehttp://www.ipedr.com/vol12/11-C011.pdf. (July 11, 2012).

Dumitru, C.O. & Gavat, I. (2006). Speaker Verification for Romanian Language. [online].

Available http://www.scientificbulletin.upb.ro/rev\_docs\_arhiva/full75923.pdf. (July 26, 2012).

Dumlao, D. (2012). Visa: foreigners’ spending in Philippines up. Philippine Daily Inquirer. [online]

available

[http://globalnation.inquirer.net/28171/visa-foreigners%E2%80%99-spending-in-](http://globalnation.inquirer.net/28171/visa-foreigners%E2%80%99-spending-in-philippines-up)

[philippines-up](http://globalnation.inquirer.net/28171/visa-foreigners%E2%80%99-spending-in-philippines-up). (March 16, 2012).

Endriga, D.A. (2010). The Dialectology of Cebuano. [online]. Available

http://mlephil.wordpress.com/2012/04/04/the-dialectology-of-cebuano/#more-5177. (July 11,

2012).

Frisbie, P. Jr. (2005). Re: What is the best format to record Speech files. [online]. Available

http://newsgroups.derkeiler.com/Archive/Comp/comp.speech.research/2005-11/msg00006.html.

(July 26, 2012).

Gupta, A. (2009). Audio Processing on Constrained Devices. [online]. Available

<http://uwspace.uwaterloo.ca/bitstream/10012/4830/1/Thesis.pdf>. (July 24, 2012).

Joseph, N. (2009). Differences Between Male and Female Speech. [online]. Available

<http://www.ehow.com/info_8089853_differences-between-male-female-speech.html>.(July 29,

2012).

Jumilla, L. (2012). Russia imports Pinoys for Apec leaders’ summit. ABS-CBN News. [online]. Available

<http://www.abs-cbnnews.com/-depth/09/05/12/russia-imports-pinoys-apec-leaders-summit>.

(October 1, 2012).

Keogh, E., & Pazzani, M. (2001). Derivative Dynamic Time Warping. [online]. Available

<http://www.ics.uci.edu/~pazzani/Publications/keogh-kdd.pdf>. (October 3, 2012).

Mandal, P. & Jha, P.K. (2012). A Comparative Study of Techniques to Implement Text-Independent

Speaker Recognition System. International Conference on Emerging Trends in Computer and Electronics Engineering. [online]. Available <http://psrcentre.org/images/extraimages/312728.pdf>. (Sept. 9, 2012).

Marking, T. (2005). Cebuano Study Notes. [online]. Available

<http://www.tomandcathymarking.com/cebuano/CebuanoStudyNotes.pdf>. (July 11, 2012).

Mosa, G., & Ali, A. (2009). Arabic Phoneme Recognition using Hierarchical Neural Fuzzy Petri Net and

LPC Feature Extraction. Signl Processing: An International Journal. [online]. Available

<http://www.cscjournals.org/csc/manuscript/Journals/SPIJ/volume3/Issue5/SPIJ-46.pdf>. (July, 18, 2012).

Mozo, M.(2012). Cebu traders pushed to do business with Australia, New Zealand. Manila Bulletin

Publishing Corporation. [online]. Available

[http://www.mb.com.ph/articles/354394/cebu-](http://www.mb.com.ph/articles/354394/cebu-traders-pushed-to-do-business-with-australia-new-zealand)

[traders-pushed-to-do-business-with-australia-new-zealand](http://www.mb.com.ph/articles/354394/cebu-traders-pushed-to-do-business-with-australia-new-zealand). (March 16, 2012).

Narayanan, R., Rajasekharan Nair, N.m & Iyappan, S. (2008, June 6). A Research Report on Engineering Students’ Performance in English Language Speaking Test. *LANGUAGE IN INDIA, Strength for Today and Bright Hope for Tomorrow*. 8(6). 2.

Navarro Jr, R. (2007). Recognition of Tagalog Alphabets Using The Hidden Markov Model.

[online]. Available [http://www.nscb.gov.ph/ncs/10thNCS/papers/invited%20papers/ips- 16/ips1601.pdf.](http://www.nscb.gov.ph/ncs/10thNCS/papers/invited%20papers/ips-16/ips16-01.pdf.) (July 11, 2012).

Oxford University Press. (2012). The OEC: Facts about the language. [online]. Available

<http://oxforddictionaries.com/words/the-oec-facts-about-the-language>. (July 18, 2012).

Friedman H. & Amoo, T. (1999). Rating the Rating Scales. [online]. Available.

<http://academic.brooklyn.cuny.edu/economic/friedman/rateratingscales.htm>. (October 14, 2012).

Wilmut, R. (2011). Principles of Audio Noise Reduction. [online]. Available <http://www.wilmut.webspace.virginmedia.com/notes/noise>. (October 14, 2012).

Eaton, S. (2009). Simple Low Pass Noise Filter. [online]. Available <http://blog.strobotics.com.au/2009/06/18/simple-low-pass-noise-filter/> (October 15, 2012).

Paulus, J.W. (2008). Advantages and Disadvantages of Wav Format. [online]. Available

<http://www.helium.com/items/967029-advantages-and-disadvantages-of-the-wav-format>. (July

24, 2012).

Rabiner, L. (1989). A Tutorial on Hidden Markov Model Models and Applications in Speech

Recognition. [online]. Available <http://www.cs.ubc.ca/~murphyk/Bayes/rabiner.pdf>. (July 29,

2012).

Saha, G., Chakroborty, S. & Senapati S. (2005). A New Silence Removal and Endpoint Detection Algorithm for Speech and Speaker Recognition Applications. [online]. Available <http://www.freewebs.com/controllab_iitkgp/ncc05.pdf>. (Aug. 13, 2012).

Salvador, S. & Chan, P. (2007). Toward Accurate Dynamic Time Warping in Linear Time and Space.

[online]. Available <http://cs.fit.edu/~pkc/papers/ida07.pdf> (September 12, 2012).

Smith III, J.O.(2011). Spectral Audio Signal Processing. [online]. Available

<https://ccrma.stanford.edu/~jos/sasp/>. (September 27, 2012).

Tiwari, G., Pandey, M. & Shrestha, M. (2011). Text-Prompted Remote Speaker Authentication. [online]. Available <http://www.scribd.com/doc/57768654/22/End-point-detection-and-Silence-removal>. (Sept. 8, 2012).

Wadhani, O., Kolhe, A., & Dekate, S. (2011). Recognition of Vernacular Language Speech for Discrete

Words using Linear Predictive Coding Technique. International Journal of Soft Computing and Engineering. [online]. Available <http://www.ijsce.org/attachments/File/Vol-1_Issue-5/E0187091511.pdf>. (July 11, 2012).

Wicklin, R. (2012). What is Mahalanobis distance? [online]. Available

<http://blogs.sas.com/content/iml/2012/02/15/what-is-mahalanobis-distance/>. (September

27, 2012).

Yu, M. (2012). Education department to use mother-tongue based in school. Sun Star Cagayan de Oro.

[online]. Available

[http://www.sunstar.com.ph/cagayan-de-oro/local-news/2012/03/26/](http://www.sunstar.com.ph/cagayan-de-oro/local-news/2012/03/26/  education-department-use-mother-tongue-based-language-school-21)

[education-department-use-mother-tongue-based-language-school-21](http://www.sunstar.com.ph/cagayan-de-oro/local-news/2012/03/26/  education-department-use-mother-tongue-based-language-school-21). (March 26, 2012).

Oregon State University. (2012). ECE 352: Signals and Systems II. [online]. Available

<http://eecs.oregonstate.edu/education/courses/ece352/>. (September 9, 2012).

“Cebuano Language– Wikipedia, the free encyclopedia” n.d. Web. 2012 Aug. 31. <<http://en.wikipedia.org/wiki/Cebuano>\_language>

“Cebu and Central Visayas – SE Asian Languages and Cultures” n.d. Web. 2012 Sept. 8. <http://www.seasite.niu.edu/Tagalog/Cebu%20Culture/CebuMain.htm>

“Recognition – Learn Cebuano” n.d. Web. 2012 Sept. 8. <http://www.101languages.net/cebuano/recognition.html>

“Introduction to Speech Processing – Oregon Health & Science University” n.d. Web. 2012 Sept. 15. <http://www. seminar.csee.ogi.edu/reu/speechsignalprocessing2.pdf>

**Video:**

Ambikairajah, E.[UNSWLearning]. (2010). Speech and Audio Processing 3: Linear Predictive Coding.

[online]. Available <http://www.youtube.com/watch?v=lWH-Oh5KnNY&feature=plcp>. (September 30, 2012).