Using Maximization Entropy in Developing a Filipino Phonetically Balanced Wordlist for a Phoneme-level Speech Recognition System

John Lorenzo Bautista*, Yoon-Joong Kim

Hanbat National University, Daejeon 305-719, South Korea

*Corresponding Author: johnlorenzobautista@gmail.com

Abstract

In this paper, a set of Filipino Phonetically Balanced Word list consisting of 250 words (FPBW250) were constructed for a phoneme-level automatic speech recognition system for the Filipino language. The Entropy Maximization Formula is used to obtain balance phonological balance in the list. Entropy of phonemes in a word is maximized, providing an optimal balance in each word's phonological distribution using the Add-Delete Method (PBW Algorithm) and is compared to the modified PBW Algorithm implemented in a dynamic algorithm approach to obtain optimization. The Filipino PBW list was extracted from 4,000 3-syllable words out of a 12,000 word dictionary and gained an entropy score of 4.2791 for the PBW Algorithm and 4.2902 for the modified algorithm. The PBW250 was recorded by 20 male and 20 female respondents, each with 2 sets data. Recordings from 30 respondents (15 male, 15 female) were trained to produce an acoustic model using a Phoneme-Based Hidden Markov Model (HMM) that were tested using recordings from 10 respondents (5 male, 5 female) using the HMM Toolkit (HTK). The results of test gave the maximum accuracy rate of 97.77% for a speaker dependent test and 89.36% for a speaker independent test.

Keywords: Entropy Maximization, Filipino Language, Hidden Markov Model, Phonetically Balanced Words, Speech Recognition

1. Introduction

Statistical models for Automatic Speech Recognition (ASR) systems are considered to be most widely used model in decoding speech into its corresponding word sequences. This model requires a large amount of speech data for training, testing, and evaluating. To provide a good

speech data for recording, the scripts for recording should represent the language as a whole. In this case, a Phonetically Balanced Wordlist (PBW) should be constructed to provide an equal balance of phonemes that represents a certain language.

There have been previous studies relating to the development of PBW in the past. In [1], a mathematical way of obtaining PBW was first introduced based on the Entropy Maximization formula, an algorithm called Add-Delete Method or simply the PBW Algorithm.

The principle of a maximum entropy states that the probability is distributed in a more balanced manner when the entropy score is at largest. The goal of PBW Algorithm is to use a greedy search algorithm to find pair of words from the initial word data list that will give an increase in entropy for a given list. However, since the algorithm presented in [1] is implemented in a greedy search approach, an optimal balance in the list is not guaranteed.

This algorithm is then used and improved in several studies such as in [2] where in an improved performance in the PBW Algorithm using Information Theory. This algorithm is called the Phonetically Optimized Wordlist (POW) algorithm.

Also, [3] proposed an efficient algorithm in selecting phonetically balanced scripts for a large-scale multilingual speech corpus. A greedy algorithm approach was applied in [3] based on distinct syllables in a word.

Furthermore, a PBW list for the Ilokano Language – a dialect in the Philippines that has similar phonemes to Filipino – was developed for human audiological examinations [4]. The word candidates used in [4] are picked based on the syllable length to ensure lesser distortion in the phonetic balance. Although this is developed for the medical field, [5] proposed a Filipino PBW produced for ASR systems based on the same algorithm presented in [4]. However, the algorithm

presented in [4] and [5] lacks a strong mathematical foundation in the providing a phonetic balance in a wordlist.

The goal of this paper is to produce a PBW list consisting of 250 words for the Filipino Language (FPBW250) based on the key ideas presented in [1][2][3] and [4] as follows:

1. Create a wordlist based on the concept of Entropy Maximization
2. Setting priorities to words with higher concentration of unique phonemes to maximize even distribution.
3. Select word candidates based on a specified syllable length to ensure lesser distortion in the phonetic balance.

Furthermore, this algorithm implemented a modified PBW Algorithm using a Dynamic Algorithm approach to ensure an optimal balance of phonemes.

Thus, we propose a modified PBW algorithm based on Information Theory and is implemented using a dynamic algorithm approach. This study is a preparation for the development of a phoneme-based large vocabulary automatic speech recognition system and N-gram based language models for the Filipino language.

This paper is organized as follows: Section 2 describes the methodology and development of the PBW250. Subsection 2.2 shows in details the source of data word entries used in this study, while Subsection 2.3 and 2.4 is about the Word Candidates and Word Selection process respectively. Subsection 2.5 shows the steps in the PBW Algorithm and the proposed Modified PBW Algorithm. Section 3 shows the methodology for testing the PBW250 using a Phoneme-based HMM recognition system based on HTK. Section IV shows the results from the testing, and finally, Section 5 is a brief conclusion and presentation of future works.

2. Development of FPBW250

2.1 Data Word Entry Source

The word entries were extracted from a medium sized tri-lingual dictionary *Diccionario Ingles-Español-Tagalog* (English-Spanish-Tagalog Dictionary) [6]. This dictionary consists of 14,651 Entries in Tagalog. Tagalog is the primary register of the Filipino language based on a dialect

spoken in Central Luzon, particularly in the Philippine's capital: Manila, and is one of the two official languages of the Philippines, other being English [7].

The dictionary includes diacritics or stress markers in its 14,651 entries as follows:

Table 1. Diacritics/Stress Markers in the Dictionary
Diccionario Ingles-Español-Tagalog

Stress	Diacritic	Example	IPA
Quick	Accute (´)	<i>Pitó</i> – seven	/pi'to/
Grave	Grave (`)	<i>Punò</i> – tree	/ˈpunoʔ/
Rushed	Circumflex (^)	<i>Punô</i> – full	/pu'noʔ/

The official spelling system for the Filipino language that uses diacritical marks for indicating long vowels and final glottal stops was introduced in 1939. Although it is used in some dictionaries and Tagalog learning materials, it has not been generally adopted by native speakers [8].

Diacritics are considered to be essential to differentiate different homophones and homographs from each other; however, there are significant differences in the recognition of spoken words by machine with reference to lexical stress [8]. Thus the word entries are narrowed down to 12,971 entries by removing diacritics.

2.2 Word Candidates

Word candidates (4,842) were selected from the 12,971 entries from the medium-sized trilingual dictionary. The word entries were selected based from the following criteria:

1. *Syllable length.* The word syllable length of the candidates is set to three (3) based on the most occurring syllable lengths of the word entries from the dictionary. Three syllable words account for 37.32% (4,842)
2. *Homophones and Homographs.* Words with the same pronunciation but different meanings (homophones) as well as words with the same spelling with variation due to lexical stress (homographs) were considered as one word candidate.

Examples:

Homophones - *mahal* (love, expensive)

Homographs - *puno* (punô - full, punò - tree)

Table 2. Syllable Length Of Word Entries

Syllable Length	Frequency
1	95
2	2520
3	4842
4	3517
5	1427
6	438
7	114
8	18
Total	12,971

2.3 Word Selection

The words were selected to form a list that should be phonetically balanced in which all the phonemes should be equally (or almost) distributed. There is no exact method to equally distribute the phoneme into the list; however, a mathematical method could be used to obtain an optimal balance of these phonemes, called *Entropy Maximization*.

$$H = - \sum_k p(k) \log_2 p(k) \quad (1)$$

Entropy H is calculated with the formula (1) where $p(k)$ is the occurrence probability of a phoneme k .

An increase in the value of entropy H would mean that the distribution of phonemes occurs almost at random. This will obtain a close to optimal balance in phonological distribution.

PBW Algorithm and a Modified PBW Algorithm

The PBW Algorithm was first introduced by employing the “Add and Delete” method [1] to maximize the value of entropy with the following procedure:

- Step 1. Add a word to the list to maximize the entropy H until the word list reaches 250 words
- Step 2. Find a pair of words that gives a maximum gain in entropy by deleting one word from the list and replacing it with the word that maximizes it.
- Step 3. Exchange the words found in step 2
- Step 4. Repeat steps 2 and 3 until there is no more gain in entropy H .

A modified algorithm similar to the Phonetically Optimized Wordlist for tri-phones in Korean [2] was

applied to the Add and Delete method to achieve a much optimal result for the phoneme distribution. A few modifications were done in the estimation of the entropy as follows:

- Step 1. Compute the number of unique phonemes per word in the candidate word list.
- Step 2. Sort the word list in descending order based on the number of unique phonemes per word.
- Step 3. Find the word in the candidate list that gives the maximum entropy value for each iteration, this will be the “maximum word”
- Step 4. Add the words into a cache list if until there is no more increase in entropy.
- Step 5. If there is no more increase in the entropy, add the words in the temporary list into the accepted list and clear the cache.
- Step 6. Continue Steps 3-5 until the accepted list reaches 250 words.

This algorithm was based on Information Theory, that the words containing the most number of phonemes will be most likely to increase the value of the entropy.

2.4 Results of the PBW Algorithm and Modified PBW Algorithm

Both the PBW Algorithm and the Modified PBW Algorithm were applied for the word candidates and obtained two different word lists. The mean frequency and standard deviation of phonemes in the word list were also computed.

The phoneme distributions for the original PBW Algorithm and the Modified PBW Algorithm are shown in Tables 3 and 4.

Table 3. Output Phoneme Distribution Table of Vowels

Vowel	PBW Algorithm Frequency	Modified Algorithm Frequency
A	189	180
E	125	127
I	148	153
O	144	145
U	144	144
Average	150	149.8
Std. Dev	21.08	17.31358

Table 4. Output Phoneme Distribution Table of Consonants

Consonant	PBW Algorithm Frequency	Modified Algorithm Frequency
CH	20	20
NG	72	75
B	78	79
D	75	77
G	77	84
H	67	72
K	79	84
L	82	83
M	83	87
N	85	87
P	77	83
R	84	86
S	81	88
T	85	84
W	57	60
Y	70	73
Average	73.25	76.375
Std. Dev	15.53	16.21

An entropy value of 4.2791 was calculated based on the PBW Algorithm and entropy of 4.2902 based on the modified PBW Algorithm can be compared in Table 5. The phonological distribution of the modified PBW Algorithm is more balanced compared to the original PBW Algorithm because of the higher entropy value. A graphical representation of the distribution of phonemes can be observed in Figures 1 and 2. An increase in the standard deviation value could also be noticed in the consonant distribution for the modified algorithm. This is because the list has already maximized the maximum frequency of the “CH” phoneme in the list. Historically, the phone CH does not appear in traditional Filipino Phoneme list [9], and thus could distort the balance in the list due to its minimal frequency. Although a higher standard deviation value is computed for the consonant list of the Modified PBW Algorithm, it wouldn’t imply that it’s less balanced than the other.

Table 5. Entropy Values Of The Pbw

	PBW Algorithm	Modified Algorithm
Total Phoneme Count	1922	1971
Entropy	4.2791	4.2902

The results gathered from the modified PBW Algorithm indicate a more balanced distribution of phonemes in the list. When used as training patterns for ASR systems, the word list extracted using this algorithm assumes that the result will provide better performance in training and recognition of phonemes.

Figure 1. Phonological Distribution Histogram of Vowels

Figure 2. Phonological Distribution Histogram of Consonants

3. Testing of PBW250 Using HTK

The Hidden Markov Model (HMM) is a stochastic sequence of underlying finite state structure which is used to model an acoustic representation of data in the development of an Automatic Speech Recognition System (ASR) [10]

A phoneme model (w) -denoted by HMM parameters (λ) - is presented with a sequence of observations (σ) to recognize a phoneme with the highest likelihood given:

$$w^* = \arg \max(w|W)P(\sigma|\lambda_w) \quad (2)$$

where:

w = phoneme, W = phoneme set

σ = observation values

λ_w = HMM model for phoneme w

In this paper, the HMM Toolkit (HTK), a toolkit for research in automatic speech recognition developed by Cambridge University [10] was used to train and develop an phoneme-based acoustic model of the Filipino Language based on the PBW250 wordlist.

3.1 Speech Data and Recording

The speech data were collected from 30 native Filipino speakers. The respondents are between the ages 18-25 years old, with no speaking ailments, and at their proper disposition during the recoding.

The speakers were grouped as training speakers (15 male, and 15 female) and testing speakers (5 male and 5 female). Each speaker were asked to recorded 2 sets of word utterances to provide a better training and testing for the ASR system. The speech data would be regarded as training data and testing data. The recorded speech data were used for both training and testing of the acoustic model developed using HTK.

The recordings were conducted in an isolated room using a unidirectional microphone (Shure SM86) connected to a computer using an audio interface (Tascam US-144mkII). A distance of approximately 5-10 centimeters between the mouth of the speaker and the microphone was maintained. A speech corpus recording tool developed by the IISPL Research Laboratory of Hanbat National University was used to collect the speech data for an easier user interface for the respondents. Each data was sampled at 16kHz at mono using a linear PCM and were saved as a waveform file format (*.wav)

3.2 Feature Specifications

The HTK tool “*HCop*” was used to extract the features from each speech data. The main parameters used in the experiment consist of 39 dimensional feature vectors from the 13 MFCC coefficient values (12 MFCC + 0th energy coefficient), derivative, and acceleration (2nd derivative). The pre-emphasis coefficient value of 0.97 is used during the feature extraction.

The coding parameters used are as follows:

TARGETKIND= MFCC_0_D_A

WINDOWSIZE=250000.0

USEHAMMING= T

PREEMCOEF= 0.97

NUMCHANS= 26

CEPLIFTER= 22

NUMCEPS= 12

3.3 HMM Phonetic Model Specifications

The data were trained with a 4,5,6, and 7-state model HMM using the Baum-welch re-estimation technique via the HTK tool “HRest”. The training was performed with multiple iterating re-estimations of the HMM parameters. A total of 20 re-estimations were done for each state-models, with the first and the last state representing a non-emitting entry and exit null states.

4. Data Preparation

The performance of the ASR is tested against two types of speakers: one which is involved in the training (dependent speakers) and the other which is only involved in the testing (independent speakers).

The recognition results are evaluated using the HTK tool HResult. The analysis tool computes for the correctly recognized word using the formula:

$$\text{Correct} = \frac{H}{N} * 100\% \quad (3)$$

Where H is the number of labels recognized and N is the total number of labels.

Accuracy is computed based on the number the insertion errors that occurred, is computed using the formula:

$$\text{Accuracy} = \frac{H - I}{N} * 100\% \quad (4)$$

Where I is the number of insertion errors.

Results from the re-estimation of the HMM parameters of each state-model groups were shown in Table 6, with the highest dependent speaker recognition rate of 97.77% for the 6-state model, and an independent speaker recognition rate of 89.36% for the 6-state model.

Table 6. Maximum Recognition Rate for each n-State model for the 20 re-estimation of the HMM models

States	Dependent-Speaker Test	Independent-Speaker Test
4	85.73	81.08
5	90.56	83.58
6	97.77	89.36
7	96.19	88.53
Average	92.56	85.64

The results imply that the 6-state model provides the acoustic model representation for the phoneme-sets used in the PBW250 wordlist. The average recognition rates of the n-state models for this study are 92.56% for the dependent-speaker test and 85.64% for the independent-speaker test.

The increase in recognition rate based on the number of re-estimations for each n-state models for the dependent and independent speaker tests were represented in a graph shown in figure 3 and 4 respectively.

Figure 3. Recognition Rate of Dependent Speaker Test for each re-estimation

Figure 4. Recognition Rate of Independent Speaker Test for each re-estimation

5. Conclusion and Future Works

The Filipino Phonetically Balanced word list of 250 words (FPBW250) was developed by using the concept of Entropy Maximization. Two 250-word lists were selected from 4,000 3-syllable words extracted from a medium-sized dictionary using the Add-Delete Method (PBW Algorithm) and a modified algorithm. Both lists were compared using the entropy scores, with values 4.2791 and 4.2902 for the

original and the modified PBW algorithm – which is based on the following: 1) entropy maximization, 2) priority of unique phonemes in a word, and 3) syllabic structure – respectively. These values suggest that the list developed using the latter is more balanced and is much appropriate for the development of a PBW speech corpus given its higher entropy value. An acoustic model was developed based on the words from the PBW250 wordlist using a phoneme-based Hidden Markov Model. The acoustic model was trained and tested using the HTK toolkit which achieved the recognition rate of 97.77% for the dependent test (based on a 6-state model) and 89.36% for the independent test (based on a 6-state model). These results suggest that the PBW250 provides a good representation of the Filipino phoneme sets based on a phoneme-based HMM acoustic model.

This study is a preparation for the development of a phoneme-based automatic speech recognition system for the Filipino language. The acoustic models used in this study will be used in developing a phoneme-based large vocabulary automatic speech recognition (LVSR) system using the Hidden Markov Model (HMM) and N-gram based language models.

References

- (1) K. Shikano: "Phonetically Balanced Word list based on information entropy", Proc. Spring Meet. Of the Acoustic Society of Japan, 1984
- (2) Y. Lim, Y. Lee: "Implementation of the POW (Phonetically Optimized Words) Algorithm for Speech Database, Proc. International Conference on Acoustics, Speech, and Signal Processing, 1995
- (3) M. Liang, R. Lyu, Y. Chiang: "An Efficient Algorithm to Select Phonetically Balanced Scripts for Constructing a Speech Corpus", International Conference on Digital Object Identifier, pp. 433-437, 2003
- (4) R. Sagon, R. Uchanski: "The Development of Ilocano Word Lists for Speech Audiometry", Philippine Journal of Otolaryngology-Head and Neck Surgery, Philippines, 2006
- (5) A. Fajardo, Y. Kim: "Development of Fillipino Phonetically-balanced Words and Test using Hidden Markov Model", Proc. International Conference on Artificial Intelligence, pp. 22-25, United States of America, July 2013

- (6) S. Calderon: “Diccionario Ingles-Español-Tagalog”, Manila, Philippines, 2012
- (7) J. Wolff: “Tagalog”, Encyclopedia of Language and Linguistics, 2006
- (8) F. De Vos: “Spelling system using diacritical marks”, Essential Tagalog Grammar, 2011
- (9) <http://wika.pbworks.com/w/page/8021671/Kasaysayan>, “Ebolusyong ng Alpabetong Filipino”, (Retrieved 2012).
- (10) S. Young: “Hidden Markov Model Toolkit: Design and Philosophy”, CUED/F-INENG/TR.152, Cambridge University Engineering Department, September 2012.