2015 8th International Conference on u- and e-Service, Science and Technology

Malay Word Pronunciation Application for Pre-School Children using Vowel Recognition

Shahrul Azmi M.Y.
Technoprenuer Incubation Center
Universiti Utara Malaysia
UUM Sintok, Malaysia
shahrulazmi@uum.edu.my

***Abstract*—In Malaysia, many researchers focus on developing independent speaker speech recognition systems that use Malay Language or *Bahasa Malaysia*. Accuracy, noise robustness and processing time are concerns when developing speech therapy systems especially for children. In this study, a Malay word pronunciation test application is developed using Spectrum Delta (SPD) features and Logistic Regression classification model in an effort to improve Malay word pronunciation for pre-school children aged between 3-6 years old. Based on the 6 vowel classification rate, vowel /i/ were found to achieved the highest classification rate of 98.33% and vowel /o/ achieved the worst with 92.29%. Overall classification rate obtained was 95.11%. Results showed that the pronunciation application can assist children to improve their Malay word pronunciation. Vowel /i/, /e/, /o/ and /u/ are often mispronounced due to pronunciation habits.**

Keywords-Malay word, pronunciation test, pre-school children

I. INTRODUCTION

Computer based speech therapy and assessment is still new in Malaysia, especially using Malay language or *Bahasa Malaysia*. In this Malaysia language, children are normally taught to spell Malay words using a combination of consonant and vowel sounds such as "KATIL" represented by syllable "KA" and "TIL". There are several studies that shows that a speech therapy system that uses vowel phonemes can be used to improve Malay word pronunciation for children. A hearing impaired person can also be trained to speak *Bahasa Malaysia* properly with a good degree of intelligibility in pronouncing given words. A high degree of standard Malay vowel recognition capability is needed in all of these systems.

Although there are many studies on Malay phoneme recognition, there is still significant work needs to be done. Most of these studies use multiple frame analysis, which is a common method employed by most researchers in the area of Speech Recognition. Accuracy, noise robustness and processing time are still concerns when developing speech therapy systems, especially for children using *Bahasa Malaysia*. The accuracy aspect involves factors such as age and gender. The size of the vocal tract of children of different gender and age varies which causes their voice to have different fundamental frequencies. This motivates this study to have an objective of developing a Malay word pronunciation test application in an effort to improve Malay word pronunciation for children.

In Malaysia, Universiti Kebangsaan Malaysia (UKM) has two computer-based speech therapy systems situated in the Clinic of Audiology and Speech Sciences. They are the Kay Elemtrics VisiPitch and IBM Speech Viewer [1]. These systems are used for voice therapy, but not used for training or articulation therapy. Furthermore, these systems use English speech therapy. There are other applications like *OLTK (Optical Logo-Therapy Kit)* [2] and *VATA (Vowel Articulation Training Aid)* [3]. These systems have limitations, and not robust enough to handle real-time identification of vowels. In 2007, Tan et. al [4] developed a Malay Speech Therapy Assistance Tool (MSTAT) which is used to assists therapists in diagnosing children for language disorder and train the children suffering from stuttering problem. It uses speech technologies consisting of speech recognition, Malay Talking Head and Malay text-to-speech system.

II. MALAY WORD PRONUNCIATION

In Malaysia itself, among the active Malaysian Universities in researching Speech Recognition are Universiti Teknologi Malaysia (UTM), Universiti Kebangsaan Malaysia (UKM), Universiti Putra Malaysia (UPM), Universiti Sains Malaysia (USM) and Multimedia University (MMU). For example, UTM did research into Malay plosives sounds and Malay numbers [5]. UTM also did a study on Malay vowels based on cepstral coefficients and fusion of Dynamic Time Warping (DTW) and Hidden Markov Model (HMM) [6, 7]. USM experimented with 200 vowel signals using wavelet de-noising approach and Probabilistic Neural Network Model [8]. UPM investigated on using Neural Networks to recognized SM digits [9]. Ting and Mark (2008) converted LPC coefficients into cepstral coefficients before being fed into a Multi-layer Perceptron with one hidden layer for training and testing [10]. The Multi-layer Perceptron was able to recognize the all speech sounds.

978-1-4673-9852-7/15 $31.00 © 2015 IEEE
DOI 10.1109/UNESST.2015.25

A. Speech Recognition Engine

The Malay Word Pronunciation Application engine is based on vowel recognition process. It starts with the data acquisition, next are filtering and pre-processing, frame selection, speech signal modelling, feature extraction and finally vowel recognition processes.

B. Data Acquisition

Data collection process was done and taken from 20 Malay pre-school children from Changlun and Jitra area in Kedah aged between 3 and 6 years old. The words "ka", "ke", "ki", "ko", "ku" and "kə" were recorded from speakers representing six vowels of /a/, /e/, /i/, /o/, /u/ and /ə/. In this study, 8000 Hz sampling frequency was used to sample the vowels and up to 10 recordings were taken per speaker depending on situation convenience.

C. Spectrum Delta Feature Extraction Method

The speech feature used in this study is Spectrum Delta developed by [11]. In this Spectral Delta approach, the differences in amplitude between each frequency band were used as features. First, the band where most of the vowel energy is situated is divided into three frequency band regions. The frequency of interest is between 1 to 2350Hz and divided into three equal band regions of 780Hz.

Determine the number of features, *i,* to be extracted from the frequency band, $BW_{SpD.}$

1) Calculate the number of frequency frames, *M,* within frequency band.

$$M = round\left(\frac{3}{2} * i\right) \quad (1)$$

2) Calculate width, *FrmB*, of a frequency frame *M*

$$FrmB = \frac{BW_{SpD}}{M} \quad (2)$$

3) Calculate individual frequency frame mean intensity, K_n, from frequency magnitude *J*. *N* is the number of frequency magnitudes within *M.*

$$K_n = \frac{1}{N} \sum_{f_n=F_{low}}^{f_n=F_{high}} J(f_n) \quad (3)$$

4) With f_n being the low and high frequency for each frequency frame. F_{delta} is the size of frame shift.

$$F_{delta} = round\left(\frac{i}{2}\right) \quad (4)$$

5) Calculate Spectral Delta features, SpD_n

$$SpD_n = K_{n+F_{delta}} - K_n \quad (5)$$

D. Logistic Regression

Many researchers have used logistic regression technique in speech recognition to classify words [12, 13]. Logistic regression is a type of category of statistical models and is commonly called generalized linear models. It can be used to predict the absence or presence of an outcome based on a set of predictor variable values. Binomial logistic regression is used when the dependent variable is a dichotomy or having 2 possible values and the independent variable can be continuous variables or categorical variables, or even both. Multinomial logistic regression (MLR) handles the case of dependent variables with more than 3 levels. The details of MLR can be found in [14], [15] and [16].

III. VOWEL CLASSIFICATION

Classifications results were based on cross validation techniques. The database is randomly divided into training and testing sets in the ratio of 7 to 3. This was done for each cross validation run where each training set will be used in training the classifier model. The other 30% of the data was treated as unseen testing inputs. A total of 10 Cross Validation tests were done and their averaged classification results were computed averaged for each classifier. Fig.1 shows that MLR did best to classify vowel /i/ with 98.33% and worst for vowel /o/ with 92.29%. Overall classification rate obtained was 95.11% using MLR.

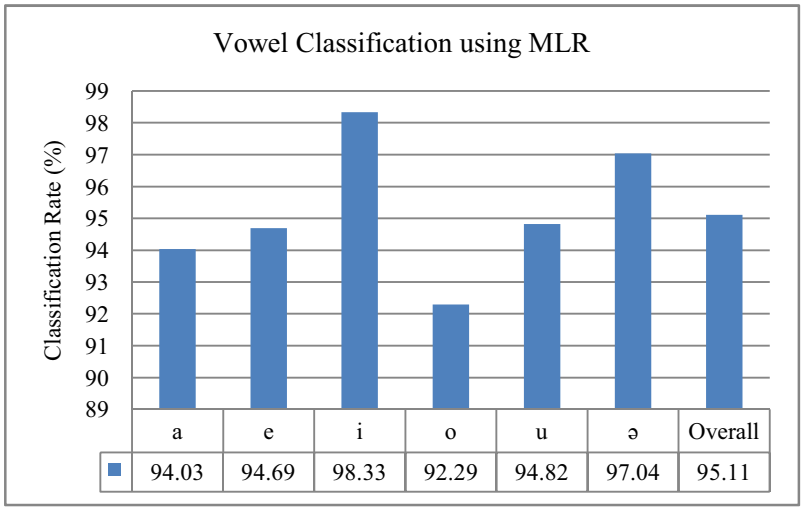

Figure 1. Result of SpD Classification Rate using MLR

IV. PRONUNCIATION TEST INTERFACE

The test interface was developed using MATLAB. The Malay words of “katil”, “roti”, “lara”, “buncis”, “potong” and “betik” were selected representing the six vowels of /a/, /e/, /i/, /o/, /u/ and /ə/. The interface is as shown in Fig. 2. The lowest accuracy was recorded based on the first uttered word in which the children are untrained. Then the children are trained on how to pronounce the words properly. Next the children will try uttering the words again 4 times and the best accuracy result is then taken.

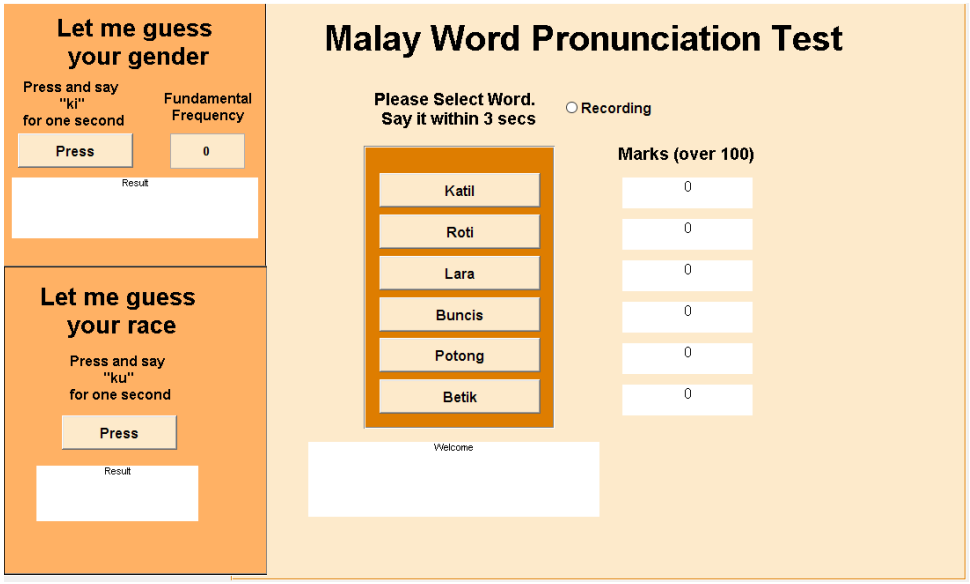

Fig. 2. MATLAB Screenshot of Testing Interface

A. Results and Discussions

The application is tested on 10 Malay pre-school children aged between 3 to 6 years from several kindergartens in Changlun and Sintok located in the state of Kedah, Malaysia. The results of the pronunciation test are shown in Table 1. The average lowest accuracy is the average accuracy of the first trial of each speaker pronunciation the given words.

TABLE 1. PRONUNCIATION TEST RESULTS

Words	Accuracy		Average Tries to reach average highest
	Average First Attempt	Average Highest of next 5 attempts	
Katil	56%	86%	4
Roti	65%	92%	4
Lara	71%	92%	2
Buncis	52%	84%	4
Potong	70%	91%	2
Betik	58%	88%	3

The first accuracy results were taken based on the first attempt. Then the speaker will be trained on how to pronounce properly and tested again in the next 5 more attempts. Then the highest score is taken from those 5 attempts. Based on the results obtained, for the word "katil" or "Ka" and "til" meaning bed, the average lowest accuracy were 56% accuracy for the first trial and 86% for after improvements were done in subsequent pronunciations. On the average, 4 times are needed to obtain the highest average accuracy. For this word, the syllable "til" was supposed to be pronounced like the word "till" in English, but often pronounced inaccurately as "tail" in English. After the correction in pronunciation, the children improve their pronunciation and obtained an improved score of 86%. For the word Roti, the children often mispronounced "Ro" as "Rue" instead of "Roo". For the word Lara, not much problem in pronouncing it correctly but the lower initial accuracy was due to the speaker spoke the word softly and sounded less confident. For the word "Buncis", the syllable "cis" was often mispronounced as "cess" instead of "cheese" in English. Not much problem seen when pronouncing the word "potong". For the word "Betik", the syllable "tik" was often mispronounced as "take" instead of "tick" in English.

V. CONCLUSION

This paper presents a new Malay Word Pronunciation Application for pre-school children. This application was developed using Matlab and uses a speech recognition algorithm based on Spectrum Delta features and logistic regression classification technique. The application was developed and tested on pre-school children aged between 3 to 6 years old. In this application, MLR classification for vowel /i/ obtained the results with 98.33% and worst for vowel /o/ with 92.29%. Based on the study, the pre-school children often mispronounced syllables having vowel /i/, /e/, /o/ and /u/. This is because of the lack of emphasizing on proper pronunciation on the given words due to daily mispronunciation which is often happening around the children. The clarity of the pronounced words may lower the accuracy measured by the application which is mostly due to nervousness and lack of confidence. Overall, this application is able to help pre-school children learn to pronounce Malay words properly and clearly.

ACKNOWLEDGMENT

The author would like to thank Universiti Utara Malaysia for providing the grant to do this study. The author would also like to appreciate all the colleagues and students who have supported in giving good insights and suggestions.

REFERENCES

[1] Ting, H., Yunus, J., Vandort, S., & Wong, L. (2003). Computer-based Malay articulation training for Malay plosives at isolated, syllable and word level. Paper presented at the Joint Conference of the Fourth International Conference on Information, Communications and Signal Processing.
[2] Hatzis, A. (1999). Optical Logo-Therapy (OLT), Computer-based audio-visual feedback using interactive visual displays for speech training.
[3] Zimmer, A. (2002). VATA: An improved personal computer-based vowel articulation training aid. Old Dominion University.
[4] Tan, T. S., Ariff, A. K., Ting, C. M., & Salleh, S. H. (2007). Application of Malay speech technology in Malay speech therapy assistance tools.
[5] M. Salam, D. Mohamad, and S. Salleh, "Neural network speaker dependent isolated Malay speech recognition system: handcrafted vs genetic algorithm," in *6th International Symposium on Signal Processing and its Applications (ISSPA2001)*, Kuala Lumpur, Malaysia, 2001.
[6] S. Al-Haddad, S. Samad, A. Hussain, K. Ishak, and A. Noor, "Robust Speech Recognition Using Fusion Techniques and Adaptive Filtering," *American Journal of Applied Sciences,* vol. 6, pp. 290-295, 2009.
[7] H. Ting and J. Yunus, "Speaker-independent Malay vowel recognition of children using multi-layer perceptron," in *IEEE Region 10 Conference (TENCON 2004)*, 2004.

[8] C. P. Lim, S. C. Woo, A. S. Loh, and R. Osman, "Speech Recognition Using Artificial Neural Networks," in *1st International Conference on Web Information Systems Engineering (WISE'00)*, Hong Kong, China 2000, p. 419.
[9] C. Tan and A. Jantan, "Digit Recognition Using Neural Networks," *Malaysian Journal of Computer Science,* vol. 17, pp. 40-54, 2004.
[10] H. N. Ting and K. M. Mark, "Speaker-dependent Malay Vowel Recognition for a Child with Articulation Disorder Using Multi-layer Perceptron," in *4th Kuala Lumpur International Conference on Biomedical Engineering 2008*, 2008, pp. 238-241.
[11] Shahrul Azmi, M., Nor Idayu, M., Roshidi, D., Yaakob, A. R., & Yaacob, S. Noise Robustness of Spectrum Delta (SpD) Features in Malay Vowel Recognition. In Computer Applications for Communication, Networking, and Digital Contents (Vol. 350, pp. 270-277): Springer Berlin Heidelberg. 2012
[12] Birkenes, Ø. (2007). A Framework for Speech Recognition using Logistic Regression. Unpublished PhD thesis, Norwegian University of Science and Technology.
[13] Hillenbrand, J., & Nearey, T. (1999). Identification of resynthesized/hVd/utterances: Effects of formant contour. The Journal of the Acoustical Society of America, 105(6), 3509-3523.
[14] Hosmer, D., & Lemeshow, S. (2004). Applied logistic regression. New York: Wiley-Interscience.
[15] Inc, S. (2003). SPSS regression models 12.0: SPSS Inc.
[16] So, Y., & Cary, N. (1995). A tutorial on logistic regression. SAS White Papers.