Sindh Univ. Res. Jour. (Sci. Ser.) Vol. 50 (3D) 210-213 (2018)

SINDH UNIVERSITY RESEARCH JOURNAL (SCIENCE SERIES)

Identification of Malay Stop Consonants Based on MFCC &Rasta PLP Features Using K-NN Classifier for Cued Speech Application

Z. ZARIDAHM[++], M. HARIHARAN*, A. K. KUSHSAIRY, M .Y .ZULKHAIRI

University Kuala Lumpur British Malaysian Institute, Batu 8 Jalan Sungai Pusu, 53100 Gombak, Selangor

Received 10th June 2018 and Revised 15th September 2018

Abstract: Phonological studies suggest that phoneme awareness in an early age through cued speech is reliable to measure the literacy skills and provide a strong language foundation for deaf children. This paper proposed a phonemic-based recognition of Malay Phonemes according to stop voicing. Eight stop consonants /p b t d tʃ dʒ k g/ preceding /a/ vowel are selected to encode each combination as a hand shape at a specified position. Features are extracted by using Mel -frequency Cepstral Coefficients (MFCC), MFCC with delta coefficients, MFCC with delta to delta coefficients and Rasta PLP. Mean of the featured group samples were taken to reduce the frame dimensions of the extracted features. These dimensionalities reduced features were fed into the k-Nearest Neighbors (k-NN) classifier for the classification. K-fold cross validation method is used to test the reliability of the classifier results. Experimental results show that the best identification rate is 92.5% upon feature fusion sets of MFCC, MFCC with delta coefficients, MFCC with delta to delta coefficients and Rasta PLP., voltage divider circuit voltage waveforms and inverter output waveforms are shown using MATLAB Simulink software.

Keywords: Cued Speech, Malay Phonemes, MFCC, Rasta PLP, Speaker Dependent

1. INTRODUCTION

Cued Speech is an alternative way of communication to the deaf, hard-hearing & language disorder children. By cueing the hearing-impaired children will have visual access to spoken language & enable them to communicate with the other normal individuals. Cued speech is not invented to replace the sign language. Cued speech uses speech reading aid containing a h& signal to indicate spoken speech sounds (Yasin, *et. al,* 2013). Cued speech enables the special children to learn a phonological model of any spoken language to facilitate them with higher learning literacy including reading, writing, speaking, &underrating. The standard configurations of cued speech use combination of eight h& gestures & four distinct locations to cue. The shapes represent consonant phonemes & the placement around the mouth represents vowel phonemes. Cued speech has been adapted to Malay phonemes & one Cued Speech Center was built in Kuala Lumpur which also known as Pusat Pertuturan Kiu.

The first publication concerning the speech intelligibility & recognition of Malay Automatic Speech Recognition (ASR) of Cued Speech was written in 2014 by Zulkhairi and Mohiuddin (Yusuf, *et. al,* 2014). They introduced a novel way to deal with measure discourse understand ability of hearing impeded youngsters by utilizing gibberish syllables as test words. The discourse clarity is estimated through the capacity of creating straightforward syllables comprising of Consonants and Vowels (CV) for 22 Malay consonants by human assessors, both guileless and master audience members. It is demonstrated that 95% of the clarity rating utilizing Malay Speech Intelligibility Test (MSIT) bought in to the comprehensibility rating of the specialists. The after-effect of Levene's Test for Equality of Variances inferred that there is prove, at the 5% level recommended that there are high likenesses in the test scores between the two assessors' gathering; language teachers and the guileless audience members. The investigations demonstrated that the MSIT is legitimate and dependable to check the discourse comprehensibility of hard of hearing Malaysian youngsters and can be for all intents and purposes utilized for hard of hearing Malaysian kids preparing and intercession program.

This paper is exhibited to help the MSIT framework with Digital Signal Processing (DSP) and Artificial Intelligence (AI) ideas. The underlying examination works began with the principal gathering of consonants as in the International Phonetics Association (IPA) outline from similar syllables' structures proposed by Zulkhairi and Mohiud commotion. Mel Frequency Cepstral Coefficients (MFCC) and Rasta-PLP were extricated as discourse highlights and the chose classifier is k-Nearest Neighbors (k-NN) classifier. The level of precision of the arrangements is exhibited in Section 4. The last area of this paper finishes up the works.

2. BACKGROUND

Phonetics is the study of a sound production & classification of speech sounds. To describe the speech

[++] Corresponding author: M.Z Zaridah. email: zaridahmz@unikl.edu.my
*School of Mechatronic Engineering, Universiti Malaysia Perlis, 02600 Arau, Perlis

sounds, phoneticians refer to a classificatory framework to describe the acoustic characteristics of a speech sound based on the voicing, place and manner of articulation as stated in the IPA chart. **(Table 1)** shows the inventory of St&red 22 Malay Consonants.

Table 1: St&ard Malay Consonants Chart

Place of Articulation Manner of Articulation	**Bilabial**	**Alveolar**	**Palatal**	**Vel ar**	**Glottal**
Plosives	p b	t d	tʃ dʒ	k g	
Nasals	m	n	ɲ	ɳ	
Fricatives	(f) (v)	s (z)	(ʃ)		h
Lateral		l			
Trill		r			
Approximants	w		y		

Stops are sometimes referred to plosive consonants, produced by forming a complete obstruction to the mouth. The plosive pairs are voiceless and voiced pairs with the voice consonants always on the right side. (**Fig 1and2)** shows the output waveforms of stop consonants of /p b t d/ & /k g c j/.

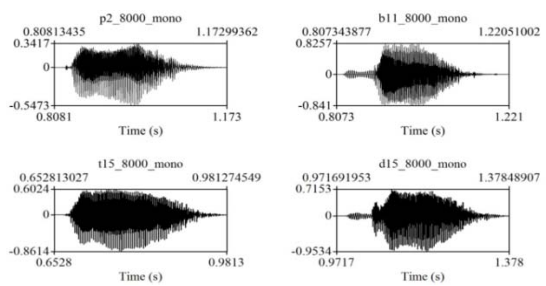

Fig.1. Output Waveforms of /p b t d/

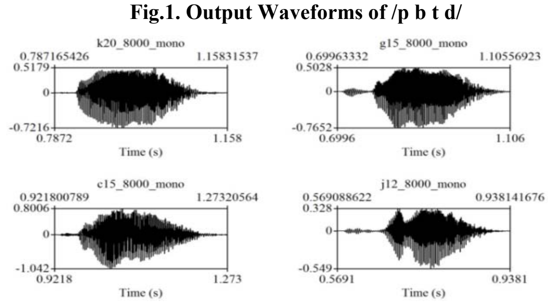

Fig.2. Output Waveforms of /k g c j/

The bilabial /b/ & /p/ (as the lips are pressed together) are having the weakest burst compared to other stop consonants. Stop consonants of /t/ & /d/ are alveolar as the tongue is pressed against the alveolar ridge. Malay Stop /t/ is not aspirated & /d/ is counterpart of /t/ where /d/ is voiced & /t/ is voiceless. Stop consonants of /k/ & /g/ are velar as the back of the tongue is pressed against an intermediate area between the hard & the soft palate. As in other st*a*red varieties, voiceless plosives are normally unaspirated since there is no flow of air out of the vocal tract during closure. Stop consonants of /tʃ/ & /dʒ/ are phonetics affricates *and* phonemically pattern with the plosive.It is detectable that the voiceless affricate/tʃ/has a frail burst taken after by an extremely solid yearning before the beginning of voicing. The goal increments in force at a considerably more noteworthy rate than the voiceless fricatives. This more quick increment in power fortifies the impression of the stop attributes of this affricate. The voiced affricate/d ʒ/shares the attributes of voiced plosives and fricatives as the burst is scarcely detectable and seems to happen at the same time with the start of the last voicing cycle before the start of the desire.

3. MATERIAL AND METHODS

Three normal Malay children are involved in the speech recording of 22 Malay Consonant Vowel (CV) syllables. The speech samples acquired at 44100 kHz sampling rate are digitized into 16-bit resolutions. The speakers uttered 30 times for each syllable. Total speech sounds recorded are 1980 samples. To proceed with the feature extraction method, speech files are segmented to remove the noise *and* silence to keep the voiced part. The following procedures are shown in **(Fig 3).**

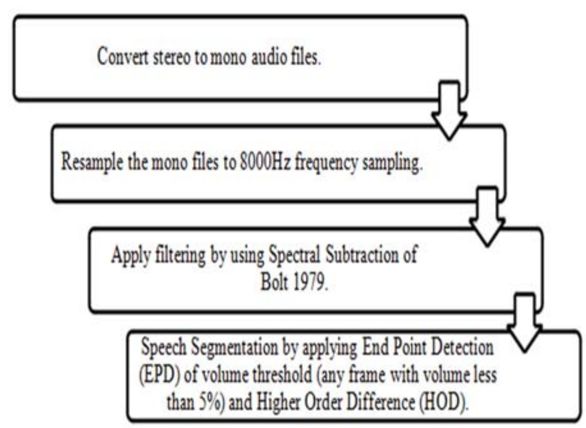

Fig.3.The Flow of Speech Segmentation

3.1 Feature Extraction

Feature extraction process converts the segmented audio signal into feature vectors which carry spectral characteristics of the speech signal. Features are extracted by Mel-frequency Cepstral Coefficients (MFCC), MFCC & delta coefficients, MFCC & delta to delta coefficients, Rasta-PLP. Feature vectors of MFCCs & Rasta-PLP are calculated using the window basis.

Mean of the featured group samples were calculated to reduce the frame dimensions of the extracted features. These dimensionalities reduced features were fed into the k-Nearest Neighbors (k-NN) classifier for the classification process. Mel Frequency Cepstral Coefficients (MFCCs) is the most widely used front-end feature extraction technique. The Mel coefficients are the logarithmic measure of the Mel magnitude spectrum, calculated by a triangular b&-pass filter. These values are decor related using Discrete Cosine Transform (DCT). MFCC is the real-valued implementation of complex cepstrum. The following steps are the generalization of the MFCC.

1. Chunk the pre-processed audio signal to short frames (100 frames/second).
2. Pre-emphasized filter with the following equation:

$S_n = X_n - aX_{n-1}(1)$

Where a = 0.97.

3. Apply Hamming windowing.
4. Apply triangular b&pass filtering.
5. Take the Discrete Cosine Transform (DCT).
6. Append the feature vectors to delta & delta-delta features
7. Take the average for each coefficient.

Delta & delta to delta coefficients are known as differential & acceleration coefficients. The MFCC feature vectors describe the power spectral envelope of a single frame, by calculating the MFCC trajectories & append them will improve the ASR performance. Appending the 13 MFCC coefficients to delta coefficients will give the feature vectors' length to the 26 coefficients. Delta to Delta coefficients are calculated by taking the difference from the delta coefficients. The final coefficients for MFCC with delta to delta coefficients will be 39 coefficients. Another option of MFCC is to discard the first column of DCT values, this will make up to 12 coefficients. The rest will be the same as applied to 13 MFCC coefficients. RASTA-PLP, an acronym for Relative Spectral Transform (RASTA) Perceptual Linear Prediction (PLP). PLP was originally proposed by Hermansky in 1991 as a way of warping spectra to minimize the differences between speakers while preserving the important speech information. RASTA is a separate technique that applies a b&-pass filter to the energy in each frequency sub-brand to smooth over short-term noise variations & to remove any constant offset resulting from static spectral coloration in the speech channel (Mustafa, *et. al,* 2016). The order of Rasta-PLP is selected like MFCC. Several researches of MFCC & Rasta-PLP have been presented on Malay speech recognition. The summary of research works on both feature extraction methods are summarized in the **(Table 2).** Based on the literature review, it has been inferred that classification of MFCC are recently improved, showing that MFCC has a good discriminating capability with strong features.

Table 2: Summary of previous works of Malay ASR

First Author, Year	**Database**	**Feature Extractio n Method**	**Classifier**	**Highest Accurac y %**
Zulkhair, *et. al,* 2003	4 speakers 22 Malay Syllables	PLP	Neural Network (NN)	76.10
Rosdi, *et. al,* 2008	21 speakers 6 Malay words	MFCC	Hidden Markov Model (HMM)	96
Seman, *et. al,* 2010	10 speakers 25 Malay words	MFCC	Multi-layer percepton (MLP)	84.73
Lim, *et. al,* 2012	40 speakers 40 Malay words	MFCC+ Euclidian Distance	GA	72.41
Zourm&, *et. al,* 2012	360 speakers 6 Malay vowels	MFCC	MLP	99.81
Nong, *et. al,* 2013	360 speakers 6 Malay vowels	MFCC	Self-Adjustable NN	90.09
Raseeda, *et. al,* 2013	35 speakers 828 Malay filled pauses	Energy boosted MFCC	MLP	77.0
Mazenan et. al,, 2014	10 Speakers (Speech Disorder) 10 Malay words	MFCC	HMM	50.0
Mazenan ,*et. al,* 2014	80 speakers 108 Malay words	MFCC	HMM	94.44
Seman, *et. al,* 2014	20 speakers 10 Malay words	MFCC	GA+ Feed Forward NN	95.00
Mustafa, et. al 2016	4 speakers 150 Malay Sentences	MFCC	HMM	74.00

3.2 Classification

k-Nearest neighbour (k-NN) algorithm is a method of classifying objects by using a database in which the data points are separated into separate classes to predict the cl assification of a new point (Chee, *et. al,* 2009). An object is classified by a majority vote of its neighbours, with the object being assigned to the class most common amongst its k-nearest neighbour where k is a positive integer. In the k-NN algorithm, the classification of a new test feature vector is determined by the class of its k-nearest neighbour located by Euclidean distance. For each training, different k values ranging between 1 to 10 are applied. For each value of k, the experiment was repeated for 10 times & for each training & testing.

4. RESULTS AND DISCUSION

(Table 3-4).shows the classification accuracy of both MFCC & Rasta-PLP, the first table classified for all 22 syllables while the following table classified for only stop syllables.

Table 3: Classification Results of k-NN Classifier for all 22 Malay Syllables

Selected Syllables	Feature Extraction Method	Average Accuracy %)
All Consonants	13 MFCC	87.68
	13 MFCC + 13 delta coefficients	87.93
	13 MFCC + 13 delta + 13 deltas to delta coefficients	87.22
	12 MFCC	86.06
	12 MFCC + 12 delta coefficients	87.26
	12 MFCC + 12 deltas + 12 delta to delta coefficients	86.26
	13 Rasta-PLP	83.03
	13 Rasta-PLP + 13 delta coefficients	82.12
	13 Rasta-PLP + 13 delta + 13 delta to delta Coeeficients	76.06

Table 4: Classification Results of k-NN Classifier for Malay Stop or Plosive Syllables

Selected Syllables	Feature Extraction Method	Average Accuracy %
Plosive Consonan ts	13 MFCC	92.50
	13 MFCC + 13 delta coefficients	92.22
	13 MFCC+ 13 delta+ 13 delta to delta coefficients	91.39
	12 MFCC	90.56
	12 MFCC + 12 delta coefficients	91.94
	12 MFCC+ 12 delta+ 12 delta to delta coefficients	91.11
	13 Rasta- PLP	85.69
	13 Rasta- PLP + 13 delta coefficients	86.94
	13 Rasta- PLP + 13 delta + 13 delta coefficients	81.80

10-fold cross validation schemes are used to prove the reliability of the classification results. In 10-fold cross validation scheme, the proposed feature vectors are divided roomy into 10 sets where nine sets are used for training and another one set is used for testing. The training process is repeated for ten times. The different number of MFCC & Rasta-PLP features have been used & the accuracy obtained are above 80% for stop syllables. The highest recognition rate is 90.25% obtained from 13 MFCC features. All 22 syllables are extracted based on MFCC and Rasta-PLP feature extraction method. The 22 syllables are trained and classified using k-NN. The recognition rate is decreased since the dimensions of input features are increased. The recognition accuracy for all syllables for MFCC & Rasta-PLP features are more than 80%. These results prove the effectiveness of both feature extraction methods & provide a high flexibility for choosing the number of coefficients to be used in classifier model.

5. CONCLUSION

The effects of using different feature extraction methods for different number of coefficients are being investigated. The classification result shows that MFCC gives the best accuracy of classifying the stop consonants for both experiments. The classification results indicate that the proposed feature extraction method may be used as a valuable tool for the recognition of stop consonants (syllables). In future, various feature selection & classification methods will be used with different manner of articulation of Malay consonants.

REFERENCES:

Chee. L. S., O. C. Ai, M. Hariharan, and S. Yaacob, (2009). "Automatic detection of prolongations *and* repetitions using LPCC", International Conference in Technical Postgraduates (TECHPOS), 1-4.

Hamzah, R., N. Jamil, and N. Seman, (2014). "Filled pause classification using energy-boosted mel-frequency cepstrum coefficients", The 8th International Conference on Robotic, Vision, Signal Processing and Power Applications, 311-319.

Hamzah, R., N. Jamil, and N. Seman, "Filled Pause Classification Using Energy-Boosted Mel-frequency Cepstrum Coefficients (MFCC)", The 8th International Conference on Robotic, Vision,

Lim, Y. C., T. S. Tan, S. H. S. Salleh, and D. K. Ling, "Application of Genetic Algorithm in unit selection for Malay speech synthesis system", Expert Systems with Applications, 39(5), 5376 - 5383.

Mustafa, M. B., and F. D. Rahman (2016)"A Two-Stage Adaptation towards Automatic Speech Recognition System for Malay-Speaking Children", World Academy of science, engineering & Technology, International J. of Computer, Electrical, Automation, Control & Information Engineering, 10(3), 451-454.

Mazenan, M. N., T. S. Tan, and L. C. Yong, (2014). "Computerized Method for Diagnosing *and* Analyzing Speech Articulation Disorder for Malay Language Speakers", Research Journal of Applied Sciences, Engineering *and* Technology, 7(24), 5278-5282.

Rosdi, F. and R. N. Ainon, (2008). "Isolated malay speech recognition using Hidden Markov Models". International Conference on Computer *and* communication Engineering, 721-725.Seman.

Yusof, Z. M., R. Hussain, and M. Ahmed, (2014). "Malay Speech Intelligibility Test (MSIT) for Deaf Malaysian Children", International J. of Integrated Eng.

Zulkhairi M. Y. and A. Mohiuddin, (2003). "Malay Phoneme Classification using Perceptual Linear Prediction (PLP) Algorithm".

Zourm, A., and H. N. Ting, (2012). "The Effect of Age on Formant Frequencies of Malay Children between 7-12" Third International Conference on Intelligent Systems, Modelling *and* Simulation (ISMS), 379-382