NOISE ROBUSTNESS OF FIRST FORMANT BANDWIDTH (F1BW) FEATURES IN MALAY VOWEL RECOGNITION

School of Computing
UUM College of Arts and Sciences
Universiti Utara Malaysia

shahrulazmi@uum.edu.my
noridayu@uum.edu.my
fad173@uum.edu.my

Sazali Yaacob

Universiti Malaysia Perlis

s.yaacob@unimap.edu.my

ABSTRACT

Applications that use vowel phonemes require a high degree of vowel recognition capability. The performance of speech recognition application under adverse noisy conditions often becomes the topic of interest among speech recognition researchers regardless of the languages in use. In Malaysia, there are an increasing number of speech recognition researchers focusing on developing independent speaker speech recognition systems that use the Malay language which is noise robust and accurate. This paper present a study of noise robust capability of an improved vowel feature extraction method called First Formant Bandwidth (F1BW). The features are extracted from both original data and noise-added data and classified using three classifiers; (i) Multinomial Logistic Regression (MLR), (ii) K-Nearest Neighbors (K-NN) and Linear Discriminant Analysis (LDA). The results show that the proposed F1BW is robust towards noise and LDA performs the best in overall vowel classification compared to MLR and K-NN in terms of robustness capability, especially with signal-to-noise (SNR) above 20dB.

Keywords: Malay vowels, spectrum envelope, speech recognition, noise robustness.

INTRODUCTION

Automatic speech recognition (ASR) has made great strides with the development of digital signal processing hardware and software. Currently, despite all these advances, machines still cannot match the performance of their human counterparts in terms of accuracy and speed, especially in the case of speaker independent speech recognition. Today, a significant portion of speech recognition research focuses on speaker independent speech recognition problems. The reasons are its wide range of applications, and limitations of available techniques of speech recognition.

Although there are studies on Malay phoneme recognition, most of them are still in infancy (Rosdi & Ainon, 2008) and use multiple frame analysis. Subsequently, more analyses focusing on the accuracy and processing time when developing speech therapy systems should be done to ensure a worth for value product is produced. Motivated by this necessity, this study is designed to develop a Malay speech recognition system in an effort to improve Malay vowel recognition. Applications that use vowel phonemes require a high degree of Standard Malay vowel recognition capability. In Malaysia, few studies have been done especially in the study of Malay vowel usage, independent speaker systems, recognition robustness and algorithm speed and accuracy.

When corrupted by low level noise, human listeners are still capable of recognizing speech because we can select and follow another speaker's voice (Devore & Shinn-Cunningham, 2003). Even at a packed football stadium, listeners can select and follow the voice of another speaker as long as the signal-to-noise ratio (SNR) is not too low. In terms of speech recognizers, most of these applications are affected by adverse environmental conditions. According to Uhl and Lieb (2001), it is important to suppress additive noise before the feature extraction stage of any speech recognizer. Invariance to background noise, channel conditions and variations of speaker and accent are among the main issues in noise robust applications (Al-Haddad, Samad, Hussain & Ishak, 2008; Huang, Acero, & Hon, 2001). Development of signal enhancement techniques as an effort to remove the noise prior to the recognition process is permissible but it may cause some alteration of the speech spectral characteristics. Consequently, the speech signal is not suitable to be used in the designed acoustic models of the recognizer hence deteriorating the performance of the recognizer (Kyriakou, Bakamidis, Dologlou & Carayannis, 2001). This justifies the needs of developing a robust speech recognizer which can be modeled using robust speech features.

This study is an effort to increase the Malay vowel recognition capability by using a new speech database that consists of words uttered by Malaysian speakers from the three major races: Malay, Chinese and Indians. This paper will present a robustness study on the First Formant Bandwidth (F1BW) method introduced by Shahrul Azmi (Shahrul Azmi, Siraj, Yaacob, Paulraj & Nazri, 2010), which is an improved formant method based on a single framed analysis of isolated utterances.

PREVIOUS RESEARCHES IN VOWEL RECOGNITION

Vowel Feature Extraction Methods

Human speech has strict hierarchical structure. It consists of sentences, which can be divided into words, and they are built by phonemes that are the basic voice construction elements. Vowels could be defined as phonemes with persistent frequency characteristics most expressed. These frequency characteristics represent a stable basis for construction of an efficient vowel recognizer. It is known from literature (Fant, 1970; Peterson & Barney, 1952; Wakita, 1977) that the spectral properties of male, females and child speech differ in a number of ways, especially in terms of average vocal tract lengths (VTL). The VTL of females is about 10% shorter compared to the VTL of males. The VTL of children is even shorter (up to 10%) than that of females.

The accents of British, American and Australian speakers can be classified by formant features such as formant frequency, bandwidth, and intensity (Yan & Vaseghi, 2003). Other formant features like amplitude and 2-dimensional Euclidean distance were also used for vowel classification (Carlson & Glass, 1992; Vuckovic & Stankovic, 2001). Formant characteristics of vowels produced by mandarin esophageal speakers were studied using the first three formant values of F1, F2, and F3 using Praat's linear predictive coding algorithm (Liu & Ng, 2009).

According to Hillenbrand and Houde (2003), the majority of vowel identification models assumed that the recognition process is driven by either the formant frequency pattern of the vowel (with or without a normalizing factor of fundamental frequency) or by the gross shape of the smoothed spectral envelope (Hillenbrand & Houde, 2003). Several other researchers have made excellent reviews of this study. The main idea underlying formant representations is the notion that the recognition of vowel identity is controlled not by the detailed shape of the spectrum but rather by the distribution of formant frequencies, mainly the three lowest formants of 1st Formant (F1), 2nd Formant (F2) and 3rd Formant (F3).

In Malaysia, research on speech recognition began in the late 1990s and has grown aggressively. Lim, Woo, Loh and Osman (2000) conducted an experiment on 200 vowel signals using the wavelet de-noising approach and the Probabilistic Neural Network Model. Salam, Mohamad and Salleh (2001) investigated Malay plosive sounds and Malay numbers while Tan and Jantan (2004) investigated Neural Networks to recognized SM digits. Another study includes Ting and Mark (2008) who converted Linear Predictive Coding (LPC) coefficients into cepstral coefficients before being fed into a Multi-layer Perceptron with one hidden layer for training and testing classifications. Yusof, Yaacob and Murugesa (2008) also studied formant difference features in classifying vowels.

Table 1 summarizes some important aspects of vowel recognition from recent literature. It addresses the issues of speaker type, frame analysis and accuracy of the recognition capability. This table shows that most of the recent researchers studied both dependent and independent speaker systems using mostly multi-frame analysis. The accuracy obtained was between 89% and 100% for the dependent speaker system and between 70% and 94% for the independent speaker and multi-framed analysis systems.

Table 1

Recent Related Literature on Vowel Recognition

Reference	Speaker Type	Frame Analysis	Accuracy %
Mohammad Nazari et al., 2008 (Nazari, Sayadiyan & Valiollahzadeh, 2008)	Independent	Multi Frame	93.9%
Ting & Mark, 2008 (Ting & Mark, 2008)	Dependent	Multi Frame	98-100%
Carvalho & Ferreira, 2008 (Carvalho & Ferreira, 2008)	Dependent	Multi Frame	89-96%
Bresolin et.al, 2007 (Bresolin, Neto & Alsina, 2007)	Independent / Dependent	Multi Frame	91.1% / 98.1%
Muralishankar & O'Shaughnessy, 2005 (Muralishankar & O' Shaughnessy, 2005) Speech, and Signal Processing (ICASSP 05)	Independent	Multi Frame	71.7%

(continued)

Reference	Speaker Type	Frame Analysis	Accuracy %
Merkx & Miles, 2005 (Merkx & Miles, 2005)	Independent	Multi Frame	91.5%
Ting & Yunus, 2004 (Ting & Yunus, 2004)	Independent	Single Frame	76.3%

From this list, the only literature that uses the independent and single-framed analysis system obtained an accuracy of only 76.3%.

In terms of robustness analysis, Luo, Soon and Yeo (2008) proposed a method to sharpen the power spectrum of the signal in both the frequency domain and the time domain by integrating simultaneous masking, forward masking and temporal integration effects into traditional mel-frequency cepstral coefficients (MFCC) that feature extraction algorithm. Yeganeh, Ahadi and Ziaei (2008) propose a set of noise-robust features based on the conventional MFCC feature extraction method which is based on a weight parameter. Rajnoha and Pollak (2007) use white noise and car noise to study the classification robustness of MFCC and Perceptual Linear Predictive (PLP) features. Gajic and Paliwal (2006) investigated how dominant-frequency information can be used in speech feature extraction to increase the robustness of automatic speech recognition against additive background noise. In Malaysia, Al-Haddad, Samad, Hussain, Ishak, and Noor, (2009), proposed an algorithm for noise cancellation by using Recursive Least Square (RLS) and pattern recognition by using a fusion method of Dynamic Time Warping (DTW) and the Hidden Markov Model (HMM). He collected Malay speech data from 60 speakers.

VOWEL RECOGNITION PROCESS

The Vowel Recognition process starts with the Data Acquisition process followed by filtering, pre-processing, frame selection, auto-regressive modelling, and feature extraction process depicted as in Figure 1. The data collection process was taken from 80 individuals consisting of both male and famale students and staff from Universiti Malaysia Perlis (UniMAP) and Universiti Utara Malaysia (UUM). As Malay is the official language for Malaysians of diverse ethnicities, the speakers were selected from among the three main races of Malays, Chinese and Indians.

The recordings were done using a conventional microphone and a laptop computer with a sampling frequency of 8000Hz. The words “ka”, “ke”, “ki”, “ko”, “ku” and “kə” were used to represent the six vowels of /a/, /e/, /i/, /o/, /u/ and /ə/ because vowels have more energy than consonants. Different combinations of consonants and vowels were tested but they yielded similar results in terms of the portion of vowels obtained. In this study, a sampling frequency of 8 kHz was used to sample the vowels and the recordings were done 2 to 4 times per speaker depending on situation convenience. The details of the data collection are listed in Table 2.

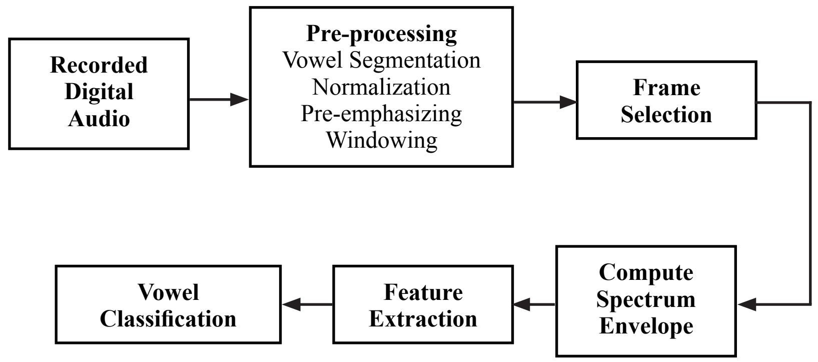

Figure 1. Vowel recognition process.

Table 2

Data Collection Detail

Information	1st Data Collection	2nd Data Collection
Sources	40 UniMAP students	20 UUM staff and 40 students
Recorded Utterances	640	728
Sampling Frequency	8000 Hz	8000 Hz
Vowels uttered	/a/, /e/, /i/, /o/, /u/, /ə/	/a/, /e/, /i/, /o/, /u/, /ə/

F1BW Feature Extraction Method

Bandwidth is the difference between the upper and lower cut-off frequencies of a signal spectrum and is measured in Hertz (Nazari et al. 2008). In signal processing, the bandwidth is the frequency at which the closed-loop system gain drops 3 decibels (dB) below peak fom equation (1) (Carlson & Class, 1992) as follows:

$$K_{BW} = \frac{1}{\sqrt{2}} K_{peak} \qquad (1)$$

Here, the K_{BW} is the resultant -3dB value that denotes the intensity value at a formant frequency, K_{peak}. Following these, two features were extracted from each vowel where the first feature was extracted based on the energy of the first formant (F1) peak (denoted by $F1BW_1$) and the second feature was extracted from the valley between the first (F1) and the second formant (F2) peaks, and is denoted by $F1BW_2$.

The steps of computing $F1BW_1$ are as follows:

i) Locate 1st formant peak ($F1_{pk}$) and its intensity ($F1_{int}$).
ii) Calculate the -3dB intensity ($BW1_{int}$) from (1).
iii) Determine the frequency range (F_{low1} < freq < F_{high1}) of $F1_{pk}$ where spectrum intensity is greater than intensity of $BW1_{int}$.
iv) Calculate mean intensity of $F1BW_1$ for each vowel using (2) where *SI* is the spectrum intensity.

$$F1BW_1(vowel) = \frac{1}{N}\sum_{f=F_{low1(vowel)}}^{f=F_{high1(vowel)}} SI(f) \tag{2}$$

Meanwhile, the steps for computing $F1BW_2$ are

i) Locate 1st formant peak ($F1_{pk}$), 2nd formant peak ($F2_{pk}$) and the valley or lowest intensity between them (FV_{low}). Their intensities are $F1_{int}$, $F2_{int}$ and FV_{int} respectively.
ii) Calculate -3dB intensity ($BW2_{int}$) value based on difference between $F2_{int}$ and V_{int} which is calculated based on:

$$BW2_{\text{int}} = F2_{pk} - \frac{(F2_{\text{int}} - FV_{\text{int}})}{\sqrt{2}} \tag{3}$$

iii) Determine the frequency range (F_{low2} < freq < F_{high2}) of FV_{int} where spectrum intensity is lower than intensity of $BW2_{int}$.
iv) Calculate mean intensity of $F1BW_2$ for each vowel using (3) where *SI* is the spectrum intensity.

$$F1BW_2(vowel) = \frac{1}{N}\sum_{f=F_{low2(vowel)}}^{f=F_{high2(vowel)}} SI(f) \tag{3}$$

In equation (3), *SI(f)* is the spectrum intensity at frequency location *f* for each vowel of /a/, /e/, /i/, /o/, /u/, /ə/ and *N* is the number of spectrum intensity value within the frequency sub-band of *BW2* for each vowel. Following these processes, the six Malay vowels represent 12 features of *F1BW1a*, *F1BW2a*, *F1BW1e*, *F1BW2e*, *F1BW1i*, *F1BW2i*, *F1BW1o*, *F1BW2o*, *F1BW1u*, *F1BW2u*, *F1BW1ə* and *F1BW2ə*.

Signal Classification Techniques

In this study, two non-linear classifiers, namely, K-Nearest Neighbours (K-NN) and Multinominal Logistic Regression (MLR), and a linear classifier based on Linear Discriminant Analysis (LDA) are used to classify all the collected features. These classifiers were chosen based on their popularities in speech recognition researches. All the computational works were conducted using MATLAB built-in functions for all the three classifiers.

F1BW Feature Analysis

Equation (1) gives twelve ranges of frequency to extract intensity features from the vocal tract model. Features of each of the vowels are extracted from two frequency bands which is the first formant peak band, and the frequency band between the first formant and second formant peak. The ranges of frequency band that are used to extract the mean intensity values from each vowel are obtained directly from the spectrum envelope of the vowels. These frequency bands are used to obtain the F1BW features.

To determine if the features of the proposed feature extraction methods significantly affect vowel classification, an ANOVA analysis was done for all the features using a statistical application called SPSS. Results of this analysis as tabulated in Table 3 show that there are significant main effects from each individual feature of the proposed F1BW method at $\alpha=0.01$ (p-value < 0.001). These results indicate that all the represented vowels, i.e. "ka", "ke", "ki", "ko", "ku" and "kə", are significantly different in each of these tested F1BW feature extraction. Therefore, the proposed extraction approach is able to show the differences of Malay spoken vowels.

Table 3

ANOVA Analysis of F1BW Features

Main Effect	df1	df2	F	Sig. (p)
$F1BW_1$	5	1310	516.42	<0.001
$F1BW_2$	5	1310	372.91	<0.001

(continued)

Main Effect	df1	df2	F	Sig. (p)
$F1BW_3$	5	1310	600.85	<0.001
$F1BW_4$	5	1310	447.16	<0.001
$F1BW_5$	5	1310	811.72	<0.001
$F1BW_6$	5	1310	461.03	<0.001
$F1BW_7$	5	1310	144.15	<0.001
$F1BW_8$	5	1310	388.92	<0.001
$F1BW_9$	5	1310	549.96	<0.001
$F1BW_{10}$	5	1310	160.41	<0.001
$F1BW_{11}$	5	1310	772.82	<0.001
$F1BW_{12}$	5	1310	478.67	<0.001

NOISE ROBUST ANALYSIS

A robust analysis was done to study the robustness of the proposed features of F1BW and to compare the results with the common single frame Mel-Frequency Cepstrum Coefficients (MFCC). White Gaussian noise was used to proof robustness. Seven signal-to-noise (SNR) levels of 10dB, 15dB, 20dB, 25dB, 30dB, 35dB and 40dB were used in this experiment in addition to the clean signal. These experiments were done on K-NN, MLR and LDA classifiers. For simplifying discussion purposes, the abbreviation "_w" refers to the classifier model which was trained with noise and "_wo" refers to the classifier model which was trained without noise. The analysis was based on cross validation testing where the original data is split randomly into 70% training set and 30% testing set (unseen input).

Two testing procedures were developed to evaluate vowel recognition performance under different training conditions. The first procedure trained 70% of all input features under different SNR including input of clean signals. This model was used to test the remaining 30% of the different SNR inputs and to measure robustness of the model when training with the features from noisy signals (see Figure 2). The second procedure trained 70% of all input features under clean input signals only. This model is then used to test the remaining 30% of the different SNR inputs including the clean features. This training method will study how robust the model is when training with the features from clean signals only (see Figure 3).

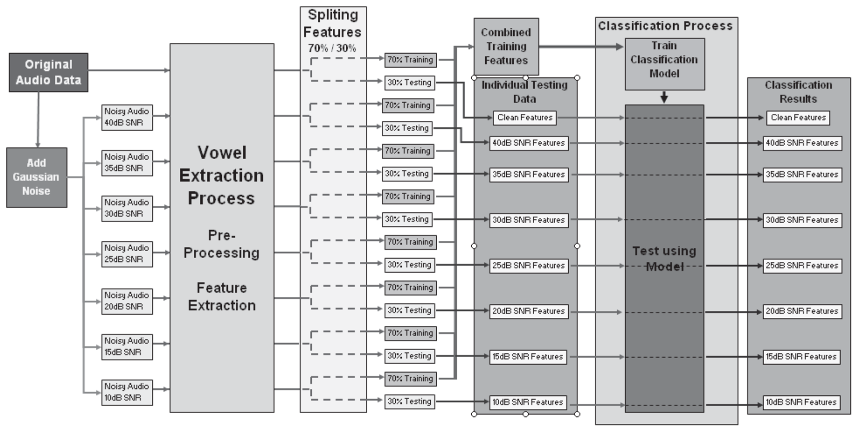

Figure 2. Robustness analysis methodology (training together with noisy data).

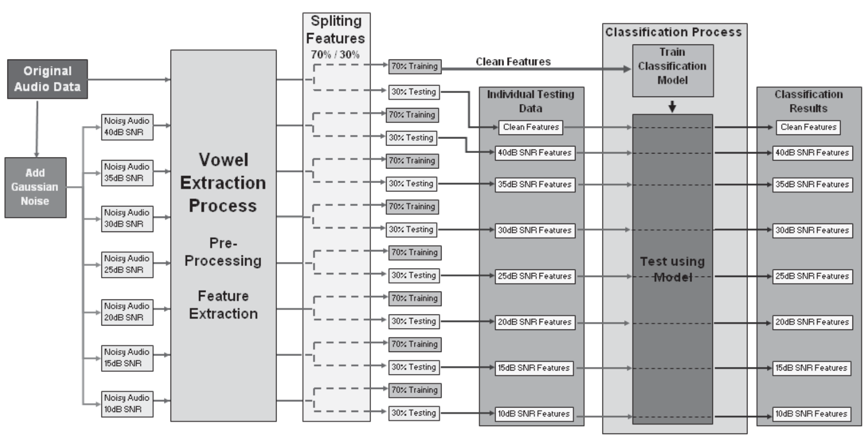

Figure 3. Robustness analysis methodology (training with only raw data).

F1BW Features Analysis

In Figure 4, the blue line represents the overall vowel classification rate of F1BW features trained with noise and tested with different SNR level data. The red line represents the overall vowel classification rate of F1BW features trained with data from raw signals only and tested with different SNR level data. For the overall vowel classification trained with only clean, the classification rate increases as SNR increases as shown by the plotted red lines in sub-Fig. 4.1. The optimum overall vowel classification rates obtained for MLR, K-NN and LDA are 93.9%, 92.5% and 90.2% respectively. For the overall vowel classification trained with noise, the MLR and K-NN overall vowel classification rates were better for SNR of 40dB and lower compared to the features trained with only clean data. As for LDA, for the overall vowel classification trained with noise, the optimum overall vowel classification rates were obtained at SNR of 30dB which is better compared to both MLR and K-NN. With regard to all classifiers, for the classification rate results trained with noisy data, “over trained” behavior was observed.

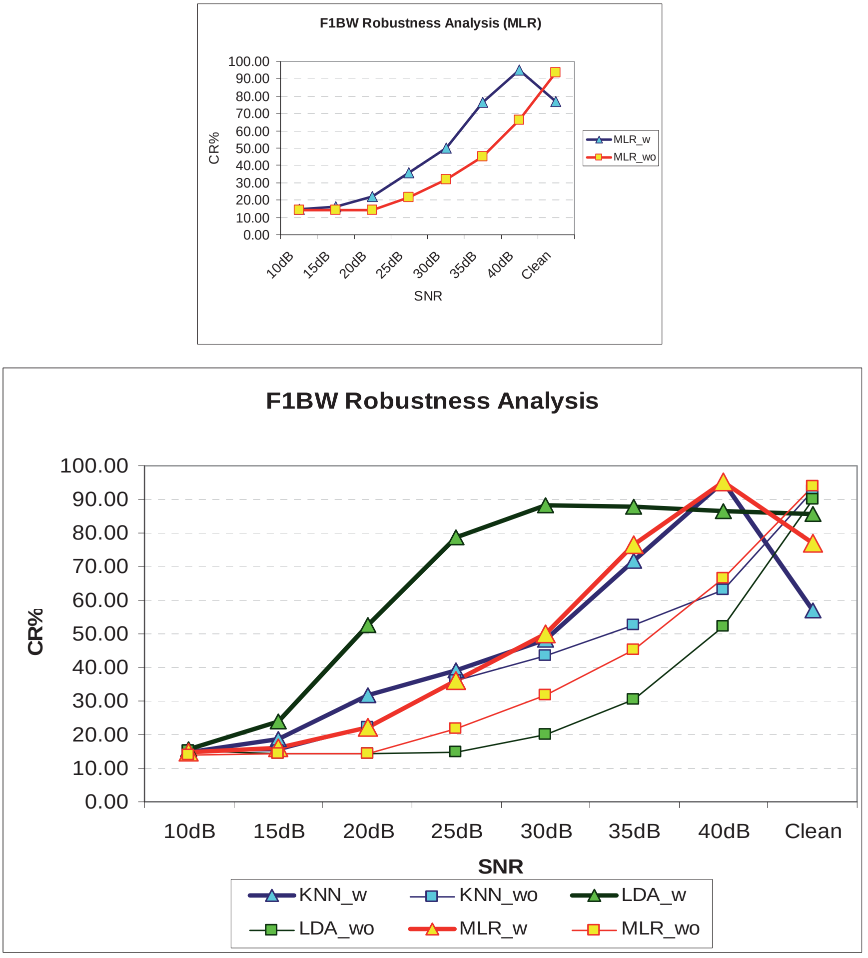

Figure 4. Overall F1BW classification rate by different SNR levels a) MLR analysis, b) KNN analysis, c) LDA analysis, d) Comparison analysis.

In Figure 4, the performance of each of the classifiers is compared. The thick colored line represents a classification model trained with noisy data and the thin colored line represents a classification model trained with only clean data. In terms of the classification rate trained with noisy data, the LDA classifier performs the best among the three classifiers because as SNR increases, the classification rate approaches the optimum faster at less than 30dB SNR which was better than MLR and K-NN, suggesting it to be the most noise robust. Furthermore, the LDA shows less "over trained" effects when compared to K-NN and MLR.

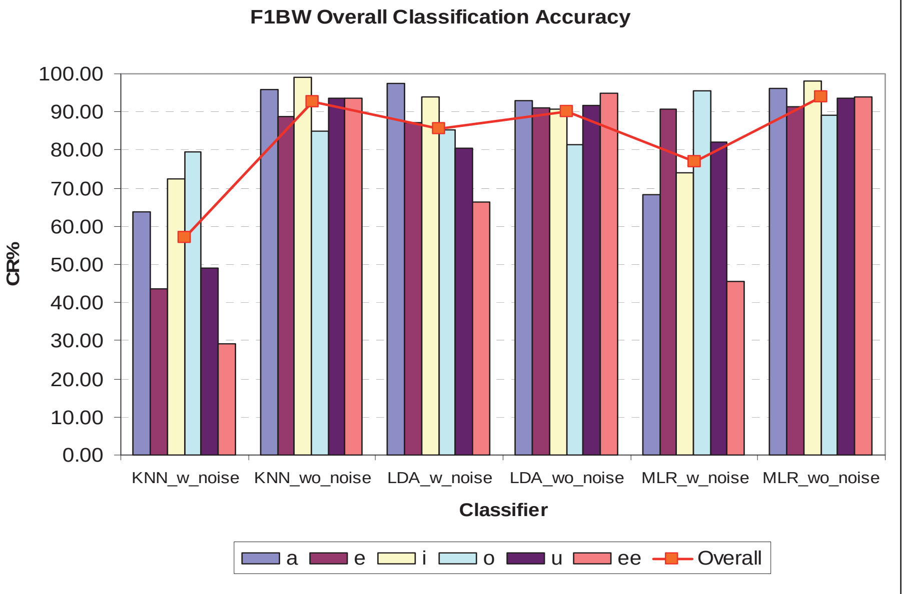

Figure 5. Overall F1BW classification rate of vowels based on classifiers and training conditions using clean training data.

Figure 5 shows the detailed overall classification result of F1BW features classified with MLR, LDA and KNN classifiers trained using only clean data. In Figure 2 and Table 1, the abbreviation "_w_noise" means that the clean trained classifier model was tested with noisy unseen data "_wo_noise" which means that the clean trained classifier model was tested with raw unseen data. Based on the overall vowel classification, the MLR classifier gave the best result of 93.8% when tested with clean data with the vowel /i/ giving the best classification accuracy as depicted in Table 4.

Table 4

Overall Classification Rate of Vowels on F1BW features using Clean Training Data (Tabulated Results)

Classifiers	Testing Data	a	e	i	o	u	ə	Overall Vowel CR%
KNN	With noise	63.7	43.5	72.6	79.4	49.1	29.0	57.1
KNN	Without noise	95.9	88.9	99.0	84.8	93.5	93.7	92.5
LDA	With noise	97.5	87.3	93.9	85.2	80.3	66.3	85.7
LDA	Without noise	92.8	91.1	90.7	81.5	91.6	94.8	90.2
MLR	With noise	68.2	90.7	74.1	95.5	82.1	45.6	77.0
MLR	Without noise	96.3	91.5	98.0	89.3	93.6	94.1	93.8

The MLR tested data with noise gives only 77.0% with /o/ giving the highest classification rate. This difference in vowel recognition performance between the classifier model trained with and without noise may be caused by how well the classifier model adapts to the noisy data. For the model which is trained with noisy data, LDA obtained the highest overall classification rate of 85.7% followed by MLR with 77.0% and K-NN with a low classification rate of only 57.1%. This makes LDA a good choice to classify vowels in a noisy environment compared to MLR and K-NN, especially where SNR is above 20dB.

CONCLUSION

This paper presents a noise robustness study on a new improved vowel feature extraction method of First Formant Bandwidth based on formant and spectrum envelope called First Formant Bandwidth (F1BW). The obtained results provide evidence that LDA is the best in overall vowel classification compared to MLR and K-NN in terms of robustness capability with less **"over-trained"** effects. It is also much better than the other two classifiers in the robustness category, especially for SNR above 20dB.

REFERENCES

Al-Haddad, S., Samad, S., Hussain, A., & Ishak, K. (2008). Isolated Malay digit recognition using pattern recognition fusion of Dynamic Time Warping and Hidden Markov Models. *American Journal of Applied Sciences, 5*(6), 714-720.

Al-Haddad, S., Samad, S., Hussain, A., Ishak, K., & Noor, A. (2009). Robust speech recognition using fusion techniques and adaptive filtering. *American Journal of Applied Sciences, 6*(2), 290-295.

Bresolin, A., Neto, A., & Alsina, P. (2007). *Brazilian vowels recognition using a New Hierarchical Decision Structure with Wavelet Packet and SVM*.

Carlson, R., & Glass, J. (1992). *Vowel classification based on analysis-by-synthesis*. Paper presented at the 2nd International Conference on Spoken Language Processing (ICSLP 92).

Carvalho, M., & Ferreira, A. (2008). *Real-time recognition of isolated vowels*. Paper presented at the Proceedings of the 4th IEEE tutorial and research workshop on Perception and Interactive Technologies for Speech-Based Systems: Perception in Multimodal Dialogue Systems.

Devore, S., & Shinn-Cunningham, B. G. (2003, 6-9 July). *Perceptual consequences of including reverberation in spatial auditory displays.* 2003 International Conference on Auditory Display, Boston, MA, USA.

Fant, G. (1970). *Acoustic theory of speech production*. Mouton De Gruyter.

Gajic, B., & Paliwal, K. K. (2006). Robust speech recognition in noisy environments based on subband spectral centroid histograms. *Audio, Speech, and Language Processing, IEEE Transactions on, 14*(2), 600-608.

Hillenbrand, J., & Houde, R. (2003). A narrow band pattern-matching model of vowel perception. *The Journal of the Acoustical Society of America, 113*, 1044-1055.

Huang, X., Acero, A., & Hon, H. (2001). *Spoken language processing: A guide to theory, algorithm, and system development*. Upper Saddle River, NJ, USA: Prentice Hall PTR.

Kyriakou, C., Bakamidis, S., Dologlou, I., & Carayannis, G. (2001, January. *Robust continuous speech recognition in the presence of coloured noise.* Proceedings of 4th European Conference on Noise Control (EURONOISE2001), Patra.

Lim, C. P., Woo, S. C., Loh, A. S., & Osman, R. (2000). *Speech recognition using artificial neural networks.* 1st International Conference on Web Information Systems Engineering (WISE'00), Hong Kong China.

Liu, H., & Ng, M. L. (2009). Formant characteristics of vowels produced by Mandarin esophageal speakers. *Journal of Voice, 23*(2), 255-260.

Luo, X., Soon, Y., & Yeo, C. K. (2008). *An auditory model for robust speech recognition.* International Conference on Audio, Language and Image Processing, 2008 (ICALIP 2008) Shanghai.

Merkx, P., & Miles, J. (2005). *Automatic vowel classification in speech* (Final Project for Math 196S). Durham, NC, USA: Department of Mathematics, Duke University.

Muralishankar, R., & O' Shaughnessy, D. (2005). *Subspace-based speaker-independent vowel recognition.* IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 05) Philadelphia, PA, USA.

Nazari, M., Sayadiyan, A., & Valiollahzadeh, S. M. (2008). *Speaker-independent vowel recognition in Persian speech.* 3rd International Conference on Information and Communication Technologies: From Theory to Applications (ICTTA 08) Umayyad Palace, Damascus, Syria.

Peterson, G., & Barney, H. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*(2), 175-184.

Rajnoha, J., & Pollak, P. (2007). *Modified feature extraction methods in robust speech recognition.* 17th International Conference of Radioelektronika, 2007, Brno.

Rosdi, F., & Ainon, R. (2008). *Isolated Malay speech recognition using Hidden Markov Models.* International Conference on Computer and Communication Engineering (ICCCE 08), Kuala Lumpur, Malaysia.

Salam, M., Mohamad, D., & Salleh, S. (2001). *Neural network speaker dependent isolated Malay speech recognition system: Handcrafted vs genetic algorithm.* 6th International Symposium on Signal Processing and its Applications (ISSPA2001). Kuala Lumpur, Malaysia.

Shahrul Azmi, M. Y., Siraj, F., Yaacob, S., Paulraj, M. P., & Nazri, A. (2010). *Improved Malay vowel feature extraction method based on first and second formants.* 2nd International Conference on Computational Intelligence, Modelling and Simulation (CIMSIM 2011). Bali, Indonesia.

Tan, C., & Jantan, A. (2004). Digit recognition using neural networks. *Malaysian Journal of Computer Science, 17*(2), 40-54.

Ting, H., & Yunus, J. (2004). *Speaker-independent Malay vowel recognition of children using multi-layer perceptron.* IEEE Region 10 Conference (TENCON 2004).

Ting, H. N., & Mark, K. M. (2008). 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* (pp. 238-241).

Uhl, C., & Lieb, M. (2001). *Experiments with an extended adaptive SVD enhancement scheme forspeech recognition in noise.* IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 01). Salt Lake City, UT, USA.

Vuckovic, V., & Stankovic, M. (2001). *Formant analysis and vowel classification methods.* 5th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Service (TELSIKS 2001).

Wakita, H. (1977). Normalization of vowels by vocal-tract length and its application to vowel identification. *IEEE Transactions on Acoustics, Speech and Signal Processing, 25*(2), 183-192.

Yan, Q., & Vaseghi, S. (2003). *Analysis, modelling and synthesis of formants of British, American and Australian accents.* IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2003).

Yeganeh, H., Ahadi, S. M., & Ziaei, A. (2008). *A new MFCC improvement method for robust ASR.* 9th International Conference on Signal Processing (ICSP 2008) Beijing, China.

Yusof, S. A. M., Yaacob, S., & Murugesa, P. (2008). Improved classification of Malaysian spoken vowels using formant differences. *Journal of ICT (JICT), 7,* Universiti Utara Malaysia.

Zahorian, S., Nossair, Z., & Norton III, C. (1999). A partitioned neural network approach for vowel classification using smoothed time/frequency features. *IEEE Transactions on Speech and Audio Processing, 7*(4), 414-425.