International Journal of Software Engineering and Its Applications
Vol.8, No.1 (2014), pp.413-426
http://dx.doi.org/10.14257/ijseia.2014.8.1.35

An Improved Feature Extraction Method for Malay Vowel Recognition based on Spectrum Delta

Shahrul Azmi M.Y.

College of Arts and Sciences, Universiti Utara Malaysia, Malaysia
shahrulazmi@uum.edu.my

Abstract

Malay speech recognition is becoming popular among Malaysian researchers. In Malaysia, more local researchers are focusing on noise robust and accurate independent speaker speech recognition systems that use Malay language. The performance of speech recognition application under adverse noisy condition often becomes the topic of interest among speech recognition researchers in any languages. This paper presents a study of noise robust capability of an improved vowel feature extraction method called Spectrum Delta (SpD). The features are extracted from both original data and noise-added data and classified using three classifiers; (i) Linear Discriminant Analysis (LDA), (ii) K-Nearest Neighbors (k-NN) and (iii) Multinomial Logistic Regression (MLR). Most of the dependent and independent speaker systems which use mostly multi-framed analysis, yielded accuracy between 89% to 100% for dependent speaker system and between 70% to 94% for an independent speaker. This study shows that SpD features obtained an accuracy of 92.42% to 95.11% using all the four classifiers on a single framed analysis which makes this result comparable to those analysed with multi-framed approach.

Keywords: *Malay Vowel, Spectrum Envelope, Speech Recognition, Noise Robustness*

1. Introduction

Speech is the most natural and a vital tool in the area human communications. Through pronunciation, thoughts and ideas are exchanged through speech. Human ability of speech production and perception marks a significant difference between man and other lower animals. Speech is important tool for social interactions in human societies. Speech processing which uses machines involves analysis and processing of speech signals that is often used for information retrieval, system control and speaker recognition. Speech processing is at the intercept of digital signal processing and natural language processing. Speech variations are affected by words, intentions, style of speaking, intonation, state of health and emotion of the speaker, accent, speaker identity, sex, gender and age. These differences are a major source of variation in speech, which can result into different pronunciations.

Although there are studies on Malay phoneme recognition, but most of them are infancy [1] and use multiple frame analysis. For example, in a speech therapy system, aspects of accuracy and processing time are of high importance. Motivated by this necessity, this study is an effort to improve Malay vowel recognition. An application that uses vowel phonemes require an accurate Standard Malay vowel recognition capability. Fortunately, there are an increasing number of studies done especially in the study of Malay vowels which focuses on independent speaker systems and recognition robustness.

When corrupted by low level noise, human listeners are still capable of recognizing speech because we can select and follow another speaker's voice [2]. Even at a noisy market situa-

ISSN: 1738-9984 IJSEIA

tion, listeners can select and follow the voice of another speaker if 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 [3], it is important to suppress additive noise before the feature extraction stage of any speech recogniser [3]. Among the main issues in noise robust applications are invariance to background noise, channel conditions and variations of speaker and accent [4, 5]. Development of signal enhancement techniques in an effort to remove the noise prior to the recognition process is permissible, but it may cause some alteration on the speech spectral characteristics. Consequently, the speech signal is unsuitable to be used in the designed acoustic models of the recognizer hence deteriorating the performance of the recognizer [6]. 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 Malay vowel recognition capability using words recorded from Malaysian speakers. This paper will present a robustness study on Spectrum Delta (SpD) method introduced by Shahrul Azmi (2010) [7] which is an improved formant method based on single framed analysis on isolated utterances.

2. Researches in Malay Speech Recognition

In Malaysia, research on speech recognition begins in late 1990s and has grown aggressively. Lim, Woo, Loh and Osman (2000) conducted an experiment on 200 vowel signals using wavelet denoising approach and Probabilistic Neural Network Model [8]. Salam, Mohamad and Salleh (2001) investigated Malay plosives sounds and Malay numbers while Tan and Jantan (2004) investigated Neural Networks to recognized SM digits [9, 10]. 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 [11]. Yusof (2008) also studied formant difference features in classifying vowels [12]. Most of the researchers in Malaysia studied on both dependent and independent speaker systems using mostly multi-framed analysis [11, 13-18]. An accuracy of between 89% to 100% was obtained using dependent speaker recorded speech and between 70% to 94% for an independent speaker speech and multi framed approach. Ting and Yunus (2004) uses an independent and single framed analysis system only obtained an accuracy of only 76.25% [18]. In terms of robustness analysis, Al-Haddad (2009), proposed an algorithm for noise cancellation by using recursive least square (RLS) and pattern recognition by using fusion method of Dynamic Time Warping (DTW) and Hidden Markov Model (HMM) [19]. He collected Malay number speech data from 60 speakers.

3. Vowel Recognition Process

Vowel Recognition process consists of data acquisition process, signal filtering, pre-processing, frame selection, auto-regressive modelling, and feature extraction process as depicted in Figure 1. Data was taken from 100 Malaysian individuals from the main three races Malays, Chinese and Indians.

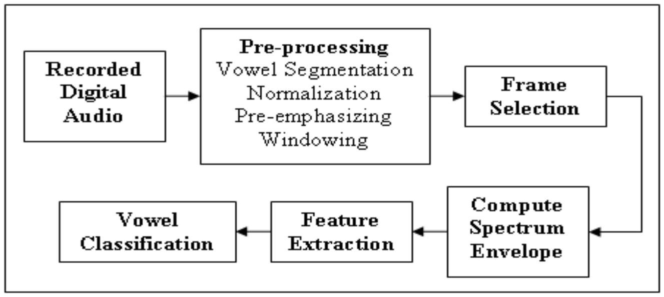

Figure 1. Vowel Recognition Process

The recordings were done using a conventional microphone and a laptop. Six vowels of /a/, /e/, /i/, /o/, /u/ and /ə/ were recorded. Different combinations of consonants and vowels were tested but yielded similar results in terms of the portion of vowel obtained. Recordings were done several times from each speakers using a sampling frequency of 8000 Hz.

3.1. Determining the Frame Size and Duration

Frame-by-frame analysis is commonly used to analyze the speech signals but a single signal frame analysis was used in this proposed vowel feature extraction method, which will reduce processing time of generating features. Previously, a Malaysian researcher has studied Malay vowel recognition using a single frame and obtains an accuracy of only 76.25% (H. Ting & Yunus, 2004). This study used vowels recorded from Malaysian children. Spectrums were analyzed using frame-shifted waveform and frame-expanding waveform methods. This is done to determine the best frame size and location to analyze on the waveform.

3.1.1. Frame Shifted

Figure 2 shows the portion of a waveform processed in the frame shifted method. The total waveform is divided into five portions, and spectrums are generated from each one of them for comparison purposes.

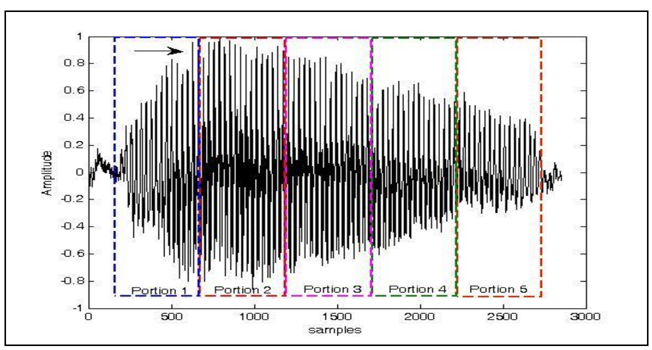

Figure 2. Frame Shifted Waveform from 0-100%

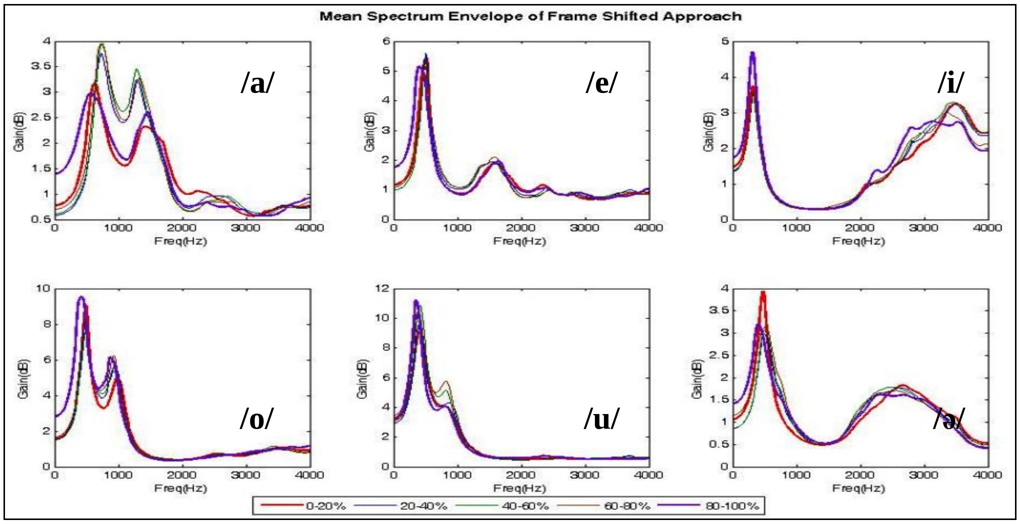

Figure 3. Frame Shifted Spectrum from 0-100%

In Figure 3, the spectrums of the frame-shifting analysis show an inconsistent response as the frame moves from left to right. The characteristics of the spectral envelope for the first 20% duration percentage are similar to the spectral characteristic of the last 20% of the duration. The middle 60% of the speech signal shows a significant variations in the spectrum. This variation may contribute to the accuracy of vowel recognition.

3.1.2. Frame Expanding Analysis

Figure 4 shows the portion of waveform processed in the frame expanding method. The total waveform is divided into ten portions, and spectrums are generated from each one of them for comparison purposes. The middle portions are situated in centre of the waveform where the most consistent characteristic of the vowels lies.

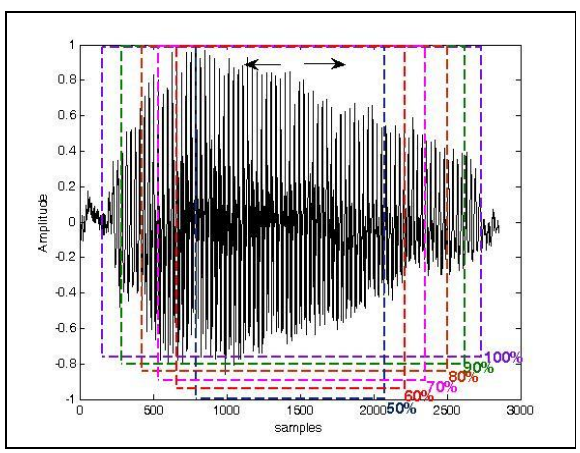

Figure 4. Frame Expanding Waveform from 0-100%

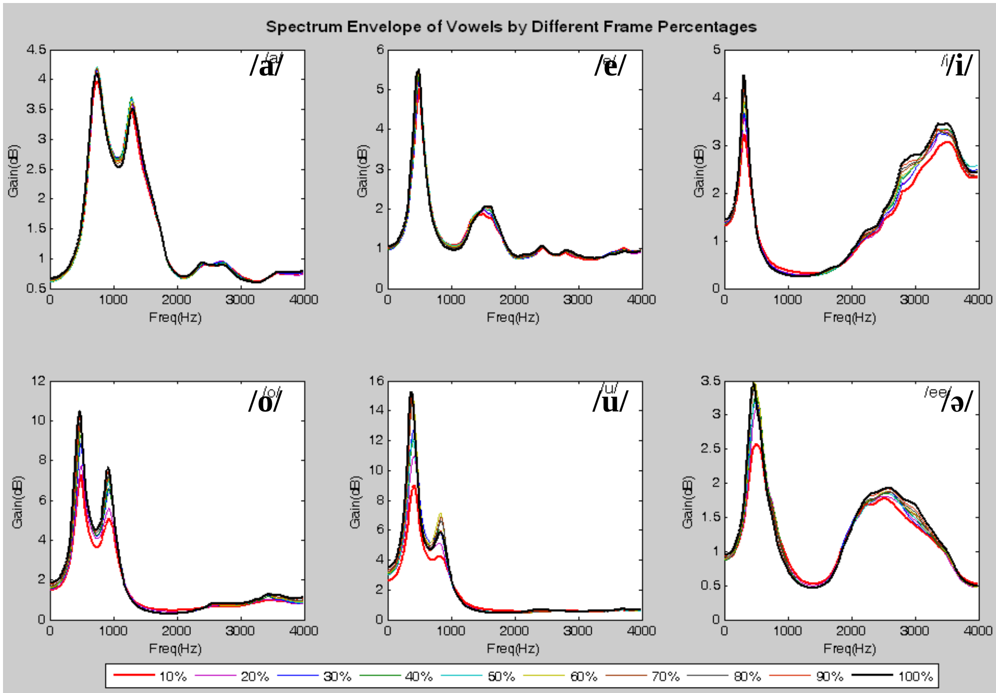

Figure 5. Frame Expanding Spectrum from 0-100%

In Figure 5, there is a significantly consistent response of the spectrums using the Frame-Expanding Analysis. In this analysis, different frame size were used with the centre of the frame being the centre of the waveform. In order to avoid initiation and termination effects, a researcher named Liu (Liu & Ng, 2009) took the medial 80% of the entire vowel, but he did not indicate how he obtained the portion size. In this study, the medial 60% waveform was chosen as the duration to be analysed.

3.2. SpD Feature Extraction Method

Spectral "Flux" (Delta Spectrum Magnitude) is the 2-norm of the frame-to-frame spectral amplitude difference vector, $\| |X_l| - |X_{l+1}| \|$. This method is somewhat similar to Hawley's method, which attempts to detect harmonic continuity in music [20, 21]. In this new Spectral Delta approach, we use the difference in Band. First, the band where most of the vowel energy is situated is divided into three regions. In this study, the frequency of interest is between 1 to 2350Hz and divided into three equal regions of 780*Hz*. The steps for SpD feature calculations are as follows:

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

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

$$M = round(\frac{3}{2} * i) \qquad (1)$$

ii. Calculate width, *FrmB*, of a frequency frame *M*

$$FrmB = \frac{BW_{SpD}}{M} \tag{2}$$

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

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

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

$$F_{delta} = round(\frac{i}{2}) \tag{4}$$

v. Calculate Spectral Delta features, SpD_n

$$SpD_n = K_{n+F_{delta}} - K_n \tag{5}$$

The visual explanation is given in Figure 6.

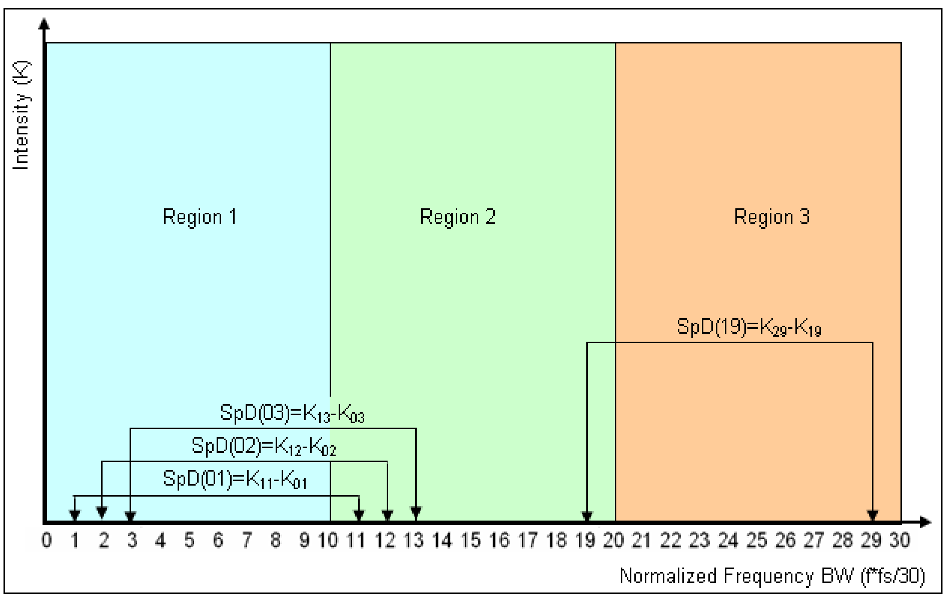

Figure 6. Visual Explanation of the Spectral Delta Calculation

Figure 6 shows three equal sized regions. Each of the normalized frequencies represents a subband of ***F***/30 Hz where ***F*** is the frequency band represented by the width from 0 to 30 in the normalized frequency scale above. SpD_{01} means the difference between mean intensity of band 11 and band 1 of the normalized frequency scale. This method uses features extracted from the intensity of harmonic discontinuity in the form of frequency bands. As explained in previous chapters, the characteristics of the spectral envelope are unique to each Malay vowel. The features reflect the rate of change between frequency bands. In order to determine the optimum number of features to extract in order to best classify the vowels, an experiment was done using a different number of features classified by Neural Network classifier. The result is shown in Figure 7.

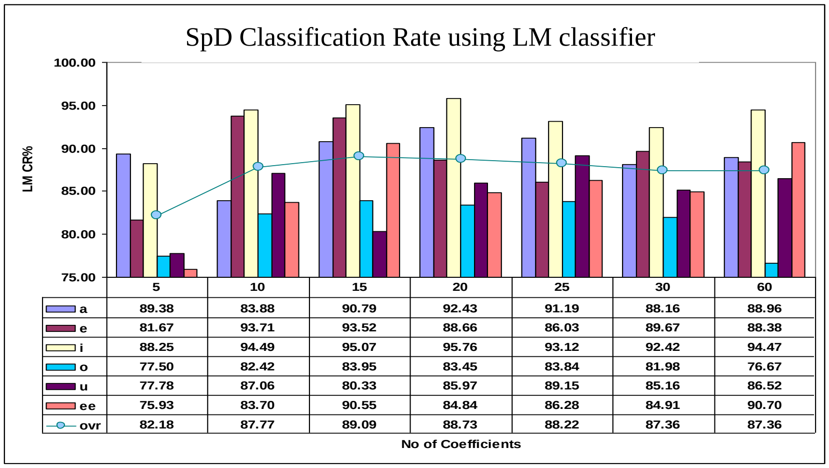

Figure 7. Spectral Delta Classification based on Number of Features

In terms of vowel performance, the classification of individual vowel is summarized in Table 1. Based on overall classification, the number of SpD features classified is 15. For the rest of the analysis of the Spectral Delta method, the number of features to be used for feature extraction is 15. The worst number of features was found to be 5 stating that this number of features does not have sufficient information to fully represent the vowels.

Table 1. Individual Vowel Classification Performance from Different SpD Number of Features

Vowel	Best Classification (Number of Features)	Worse Classification (Number of Features)
/a/	20	10
/e/	10, **15**	5
/i/	**15**, 20	5
/o/	**15**, 20, 25	5, 60
/u/	25	5
/ə/	**15**, 60	5

3.3. Vowel Classification Techniques

K-Nearest Neighbours (k-NN) and Multinomial Logistic Regression (MLR) were the two non-linear classifier used in this study. Linear Discriminant Analysis (LDA) is another linear based classifier used to classify all the collected features. Due to their popularities in speech recognition researches, these classifiers were chosen. All the computational works were conducted using *MATLAB* built-in functions for all the three classifiers.

4. Feature Analysis

This method uses features extracted from the intensity of harmonic discontinuity in the form of frequency bands. The features reflect the rate of change between frequency bands. In order to determine the optimum number of features to extract in order to best classify the vowels, an experiment was done using different number of features classified by Neural Network classifier.

Table 2. Anova Analysis of SpD Features

Main Effect	df1	df2	F	Sig. (p)
SpD_1	5	1310	173	<0.001
SpD_2	5	1310	212	<0.001
SpD_3	5	1310	226	<0.001
SpD_4	5	1310	166	<0.001
SpD_5	5	1310	54	<0.001
SpD_6	5	1310	53	<0.001
SpD_7	5	1310	231	<0.001
SpD_8	5	1310	99	<0.001
SpD_9	5	1310	36	<0.001
SpD_{10}	5	1310	39	<0.001
SpD_{11}	5	1310	102	<0.001
SpD_{12}	5	1310	119	<0.001
SpD_{13}	5	1310	38	<0.001
SpD_{14}	5	1310	8	<0.001
SpD_{15}	5	1310	9	<0.001

An ANOVA analysis was done to determine if the features of the proposed feature extraction methods significantly affect vowel classification for all the features using a *SPSS* application. Results of this analysis as tabulated in Table 2 shows that there are significant main effects from each individual feature of the proposed SpD method at α=0.01 (p-value < 0.001). These results indicate that all the represent vowels are significantly different in each of these tested SpD features extraction. Therefore, the proposed extraction approach is able to show the differences of Malay spoken vowels.

4.1. 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 20 Cross Validations tests were done and their averaged classification results were computed averaged for each classifier. Below are the steps of the 10-fold cross validation process. *N* is the total runs of validation runs.

Cross validation steps for LDA, MLR and KNN classifiers for *N* runs

Step 1: Randomize the data (total 1368)
Step 2: Split data into 2 sets
  Set 1 (70%) – Training Set (958 data)
  Set 2 (30%) – Testing Set (410 data)
Step 3: Train each of the 4 classifier models using same training set.
Step 4: Test each model with the same testing set. Compute classification rate.

Step 5: Repeat step 1-4 for next run until *N*-1 runs.
Step 6: Compute average classification rate. End

Cross validation steps for Levenberg-Marquardt (LM) network for *N* runs
Step 1: Normalize data from 0.1 – 0.9
Step 2: Randomize data
Step 3: Split data into 2 sets
Set 1 (70%) – Training Set
Set 2 (30%) – Testing Set
Step 4: Train each classifier model using same training set.
Step 5: Test model with testing set. Compute Classification rate.
Step 6: Repeat step 1-4 for next run until *N*-1 runs.
Step 7: Compute average classification rate. End

Classification of SpD features was done using 4 classifiers of KNN, MLR, LM and LDA. The result is shown in Figure 8 and Table 3.

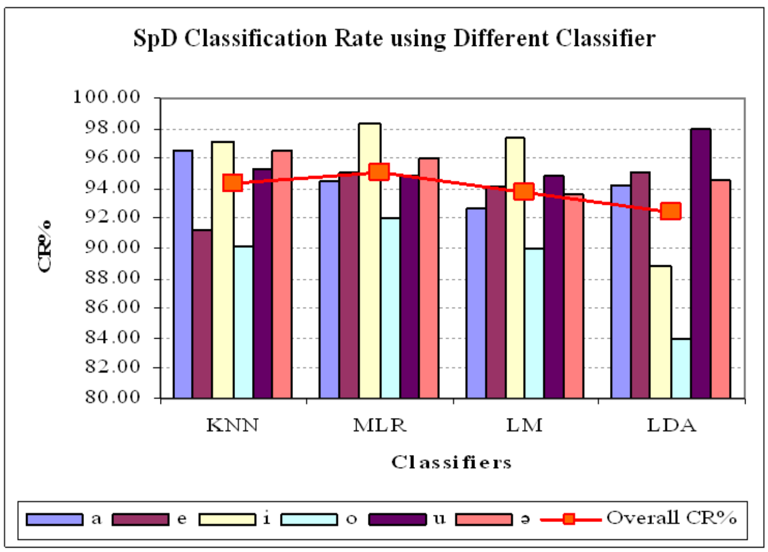

Figure 8. Result of SpD Classification Rate using Multiple Classifiers

Table 3. Classification Rate of SpD Features using Different Classifiers

Method	a	e	i	o	u	ə	Overall	Training Time (s)
KNN	96.50	91.23	97.13	90.16	95.31	96.45	94.35	0.15
MLR	94.43	95.09	98.33	91.99	94.82	96.04	95.11	9.66
LM	92.67	94.12	97.32	90.00	94.77	93.67	93.77	16.05
LDA	94.27	95.07	88.85	84.00	97.99	94.56	92.42	0.01

Table 4 shows that vowel /i/ was best classified by all classifier except LDA. MLR did best to classify vowel /i/ with 98.33% followed by LM, 97.32% and KNN of 97.13. LDA did best for vowel /u/ giving 97.99%. Vowel /o/ was classified the worst but still obtained greater than 90.00% accuracy with LDA getting 84.00%. In terms of training time, both KNN and LDA took less than 0.2 seconds to train the model compared to MLR and LM which took more than 9 seconds. MLR again gave the best overall classification rate of 95.11% which is 0.76% better than KNN. Overall, a good classification rate of above 92% was obtained for all the classifiers based on SpD features. This result proves that SpD features can be used to accurately classify vowels.

Table 4. Best and Worst Vowel Classification Result for SpD Features

	Best Recognition Performance for Vowel		Worst Recognition Performance for Vowel	
Classifier	**Vowel**	**CR%**	**Vowel**	**CR%**
KNN	/i/	97.13	/o/	90.16
MLR	/i/	98.33	/o/	91.99
LM	/i/	97.32	/o/	90.00
LDA	/u/	97.99	/o/	84.00

4.2. Noise Robust Analysis

A robust analysis was done to study the robustness of the proposed features of SpD 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 classifier model, which was trained with noise and "_wo" refers to 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).

Table 5. Comparison of Overall SpD Classification Rate by Different snr Level (Tabulated Result)

SNR	KNN_w_noise	KNN_wo_noise	LDA_w_noise	LDA_wo_noise	MLR_w_noise	MLR_wo_noise
10dB	19.90	20.03	21.13	14.86	13.81	18.02
15dB	25.96	26.83	54.58	15.93	13.63	24.90
20dB	37.30	38.92	82.80	23.30	19.57	36.37
25dB	58.76	54.56	89.00	37.41	35.41	50.86
30dB	71.90	62.80	89.37	49.19	55.41	58.38
35dB	89.23	70.51	89.70	57.69	78.25	71.15
40dB	95.18	78.44	89.64	78.27	95.94	86.28
Clean	83.82	94.77	89.37	92.41	77.18	95.09

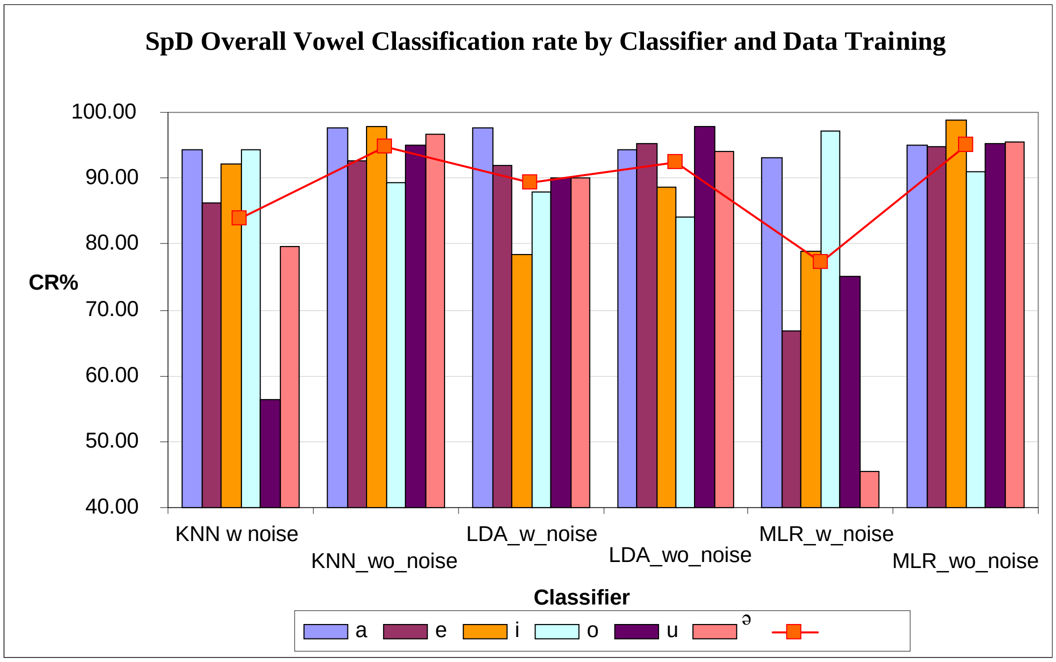

Figure 9. Overall SpD Classification Rate of Vowels based on Classifiers and Training Conditions using Clean Training Data

Figure 9 shows the detailed overall classification result of SpD features classified with MLR, LDA and KNN classifiers. It also shows the vowel recognition performance of individual vowels for all classifiers trained either with noisy data or clean data. Table 6 shows MLR_wo_noise performs the best by giving 95.09% overall classification rate for training clean data with vowel /i/ giving the highest result and the rest of the vowel achieving above 91% classification rate. MLR_w_noise gave only 77.18% for testing clean data with /o/ giving the highest classification rate. This difference in vowel recognition performance may be caused by the adaptation of the model to the noisy data. For the model which is trained with noisy data, LDA obtained the highest overall classification rate of 89.37% followed by KNN with 83.82% and KNN with a low classification rate of only 77.18%.

Table 6. Overall Classification Rate of Vowels on SpD Features using Clean Training Data (Tabulated Results)

Classifiers	a	e	i	o	u	ə	Overall
KNN w noise	94.37	86.34	92.08	94.38	56.37	79.62	83.82
KNN_wo_noise	97.70	92.62	97.83	89.38	95.11	96.77	94.77
LDA_w_noise	97.69	91.86	78.53	87.84	90.14	90.08	89.37
LDA_wo_noise	94.37	95.29	88.58	84.20	97.90	94.00	92.41
MLR_w_noise	93.11	66.77	78.78	97.21	75.10	45.52	77.18
MLR_wo_noise	95.06	94.85	98.70	91.07	95.36	95.61	95.09

5. Conclusion

This paper presents a study of noise robustness on a new improved vowel feature extraction method based on frame-to-frame spectral amplitude difference vector called Spectrum Delta (SpD). The obtained results provided evidence that LDA is the best classifier in overall vowel classification. Its performance is better than MLR and k-NN in terms of robustness capability for SNR above 20dB. As mentioned before, most of the recent researchers studied on both dependent and independent speaker systems using mostly multi-framed analysis, which yielded accuracy between 89% to 100% for dependent speaker system and between 70% to 94% for an independent speaker. This study shows that SpD features obtained an accuracy of 92.42% to 95.11% using all the four classifiers on a single framed analysis which makes this result comparable to those analysed with multi-framed approach. Single framed approach can definitely reduce processing time vowel recognition. In terms of noise robust analysis, the proposed technique of SpD yielded an average of 95.09% when trained with noise and classified using MLR, which is the best classifier among the four classifiers tested. When trained without noise, the accuracy obtained was 89.37% using LDA. This result shows that the proposed method which uses independent speaker data can yield good accuracy compared with previous studies.

Acknowledgements

This paper is a revised and expanded version of a paper entitled “Noise Robustness of Spectrum Delta (SpD) Features in Malay Vowel Recognition” presented at the International Conference on Digital Contents and Applications (DCA 2012), held in December of 2012, located at Korea Woman Training Center, Kangwondo in Seoul, Korea. I would like to thank Universiti Utara Malaysia for funding this study. I would also thank the UUM Research and Innovation Centre (RIMC) and their excellent staffs for providing support and assistance to me throughout this study. Appreciation is also given to Tadika Tunku Puteri Intan Shafinaz for providing the valuable speech data.

References

[1] F. Rosdi and R. Ainon, “Isolated malay speech recognition using Hidden Markov Models”, International Conference on Computer and Communication Engineering (ICCCE 08), Kuala Lumpur, Malaysia, **(2008)**, pp. 721-725.

[2] S. Devore and B. G. Shinn-Cunningham, “Perceptual consequences of including reverberation in spatial auditory displays”, 2003 International Conference on Auditory Display, Boston, MA, USA, **(2003)**, pp. 75-78.

[3] C. Uhl and M. Lieb, “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, **(2001)**, pp. 281-284.

[4] S. Al-Haddad, S. Samad, A. Hussain and K. Ishak, “Isolated Malay Digit Recognition Using Pattern Recognition Fusion of Dynamic Time Warping and Hidden Markov Models”, American Journal of Applied Sciences, vol. 5, **(2008)**, pp. 714-720.

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

[6] C. Kyriakou, S. Bakamidis, I. Dologlou and G. Carayannis, “Robust Continuous Speech Recognition in the Presence of Coloured Noise”, Proceedings of 4th European Conference on Noise Control (EURONOISE2001), Patra, **(2001)**, pp. 702-705.

[7] M. Y. Shahrul Azmi, “Feature Extraction and Classification of Malay Speech Vowels”, School of Mechatronics, PhD, Kangar, Perlis, MALAYSIA: Universiti Malaysia Perlis (UniMAP), **(2010)**.

[8] C. P. Lim, S. C. Woo, A. S. Loh and R. Osman, “Speech Recognition Using Artificial Neural Networks”, 1st International Conference on Web Information Systems Engineering (WISE'00), Hong Kong, China, **(2000)**, pp. 419.

[9] M. Salam, D. Mohamad and S. Salleh, "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, **(2001)**.
[10] C. Tan and A. Jantan, "Digit Recognition Using Neural Networks", Malaysian Journal of Computer Science, vol. 17, **(2004)**, pp. 40-54.
[11] H. N. Ting and K. M. Mark, "Speaker-dependent Malay Vowel Recognition for a Child with Articulation Disorder Using Multi-layer Perceptron", 4th Kuala Lumpur International Conference on Biomedical Engineering 2008, **(2008)**, pp. 238-241.
[12] S. A. M. Yusof, S. Yaacob and P. Murugesa, "Improved Classification of Malaysian Spoken Vowels using Formant Differences", Journal of ICT (JICT), Universiti Utara Malaysia, vol. 7, **(2008)** December.
[13] M. Nazari, A. Sayadiyan and S. M. Valiollahzadeh, "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, **(2008)**, pp. 1-5.
[14] M. Carvalho and A. Ferreira, "Real-Time Recognition of Isolated Vowels", Lecture Notes in Computer Science, vol. 5078, **(2008)**, pp. 156-167.
[15] A. Bresolin, A. Neto and P. Alsina, "Brazilian Vowels Recognition using a New Hierarchical Decision Structure with Wavelet Packet and SVM", **(2007)**.
[16] R. Muralishankar, L. N. Kaushik and A. G. Ramakrishnan, "Time-scaling of speech and music using independent subspace analysis", International Conference on Signal Processing and Communications (SPCOM 2004), **(2004)**, pp. 310-314.
[17] P. Merkx and J. Miles, "Automatic Vowel Classification in Speech", Department of Mathematics, Duke University, Durham, NC, USA, Final Project for Math 196S2005.
[18] H. Ting and J. Yunus, "Speaker-independent Malay vowel recognition of children using multi-layer perceptron", IEEE Region 10 Conference (TENCON 2004), **(2004)**.
[19] 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, **(2009)**, pp. 290-295.
[20] M. Hawley, "Structure out of Sound", School of Architecture and Planning. vol. PhD Massachusetts, USA: Massachusetts Institute of Technology, **(1993)**, pp. 185.
[21] E. Scheirer and M. Slaney, "Construction and evaluation of a robust multifeature speech/music discriminator", IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 97), Munich, Germany, **(1997)**, pp. 1331-1334.

Author

Shahrul Azmi M.Y., PhD, he is a Senior Member of IEEE who obtained his BSc. in Electrical Engineering from Missouri University of Science and Technology in 1995. He then worked at Seagate and Robert Bosch as an Engineer from 1995 until 1999. His research interests are speech recognition, artificial intelligence, pattern recognition and production process Control. In 2000, he joined the faculty member of Universiti Utara Malaysia (UUM) under the School of Information Technology. In UUM, he was the Deputy Dean of Academics in 2003 and currently the program coordinator for Master of Science in Technopreneurship. As an academician, he has published more than 51 articles in proceedings, journals and book chapters. He has obtained and completed 6 research grants. He is currently active in 6 researches mostly in the area of speech recognition. He has participated in various international product exhibitions and won many medals in the areas of ICT and Engineering.