Statistical Band Selection for Descriptors of MBSE and MFCC-based Features for Accent Classification of Malaysian English

M. A. Yusnita, *Student Member*, *IEEE*, M. P. Paulraj, *Member*, *IEEE*, Y. Sazali, *Member*, *IEEE*, A. B. Shahriman and M. Nor Fadzilah

***Abstract*—Accent is a major cause of speech variability that complicates the speech technology systems. Interestingly, ethnicity is one of the influential factor that give rise to accentuation in speech. Proper approach of extracting ethnical accent information is utmost crucial in many speech applications. This paper proposes an efficient way of analyzing the ethnical accent using statistical knowledge of log-energies of fourier transformed derived mel-filter banks. A simple algorithm to select bands so called statistical band selection (SBS) method using smallest variances within class scores was developed to optimize the presentation of speech features. The experiments were conducted on selective accent-sensitive words of male and female speakers originate from three major ethnics in Malaysia. Firstly, statistical descriptors such as mean, standard deviation, kurtosis and the ratio of standard deviation to kurtosis of mel-bands spectral energy and secondly, mel-frequency cepstral coefficients were extracted from the selected bands to model an accent classifier, implemented based on neural network model and K-nearest neighbors. Experimental results showed that SBS has increased the performance of accent classification system by achieving better accuracy rates between 4% to 6%, lesser memory requirement between 22% to 55% and faster speed of 70% on average of three-class accent problem.**

***Index Terms*—Statistical band selection, Mel-bands spectral energy, Mel-frequency cepstral coefficients, Accent classification, Artificial neural network, K-nearest neighbors, Malaysian English.**

Manuscript received March 10, 2013. This work was supported by the Malaysian Ministry of Higher Education and Universiti Teknologi MARA under the sponsorship of SLAB doctorate program.

Yusnita M A is with Faculty of Electrical Engineering, Universiti Teknologi MARA, Pulau Pinang, Malaysia and a Ph.D candidate at School of Mechatronic Engineering, Universiti Malaysia Perlis, Malaysia. (e-mail: yusnita082@ppinang.uitm.edu.my).

Paulraj M P, Sazali Yaacob and Shahriman A B are with School of Mechatronic Engineering, Universiti Malaysia Perlis, Malaysia. (e-mail: paul@unimap.edu.my, s.yaacob@unimap.edu.my and shahriman@unimap.edu.my).

Nor Fadzilah Mokhtar is with Faculty of Electrical Engineering, Universiti Teknologi MARA, Pulau Pinang, Malaysia. (e-mail: norfadzilah105@ppinang.uitm.edu.my).

I. Introduction

Despite recognizing speech linguistic message and speaker identity, other important traits of human voice biometrics, such as accent, gender, emotion and health state can be mined from the speech signal. Even non-experts can distinguish these traits by their common sense. Teaching machines about these knowledge has found useful applications in many speech technology systems. For example, to establish rapport and to foreground familiarity with its human partner, an intelligent humanoid should speak like one. Intrinsically, accent, defined as a systematic variation in pronunciation patterns due to different ethno-linguistic and cultural background of a speaker is a major cause of speech variability. The performance of interactive voice response systems for speaker-independent environment using telephone networks for instance, degrades when presented with accented speech of users from different geographical regions in the world. English language, although globally accepted as the most popular way of communicating, it is spoken with different flavors

of accents according to different factors. Those widely known [1] factors are such as i) ethnic, ii) age, iii) gender, iv) geographical origin, v) educational background, vi) language proficiency and usage, vii) social-economic class and viii) experience of staying in English-speaking country. In a scenario of multi-ethnic, multi-cultural and multi-lingual society composed by 28.334 million population of 50.1% Malays, 22.6% Chinese, 6.7% Indians and others [2] in Malaysia, ethnicity is the main concern that give rise to accentuation in speech. Hence proper approach and method of extracting ethnical accent information is utmost crucial for the success of accent recognizers to assist speaker and speech recognition systems (ASRs).

Malaysia English (MalE) has been accepted as a localized accent used as second most important lingua franca after the national Malay language with distinct features differs largely from the Standard English in many aspects of phonology. It evolves from the British English and has been influenced by the American English as well as the local languages such as the Malay, Mandarin and Tamil [3]. Ethnically diverse MalE accented speech sometimes termed as ethnolect by the linguists was greatly discussed in some linguistic literature such as [4-6] as these sub-variants mark the identities and cultural background. Furthermore the ethnic identity markers are useful to design an English instructor computer program that can educates users to articulate depending on major problem faced by their first languange. Research work on what features differentiate these ethnic groups speech and how to quantify the differences are rare. However research findings closer to home [7] which were conducted on Singaporean English (majority are Chinese, Malay and Indian in accordance) using human respondents proved that young Singaporeans of different ethnic groups could be identifiable from their speech. Others related studies [8-10] include vowels and intonation patterns acostical analysis.

Engineering approaches can assist this research by analyzing features associated to discriminating ethnic identity quantitatively through digital signal processing techniques and perform the recognition automatically. Numerous studies in accent recognition have attempted to use spectral analysis to extract accent features such as in filter banks analysis, mel-frequency cepstral coefficient (MFCC), perceptually linear predictive [11-15], parametric autoregressive model i.e. linear predictive coding and formant analysis [16-19]. Others employed temporal features such as pitch contour and energy [17, 20]. Filter bank analysis gives the most fundamental concept and the earliest in speech processing formulated based on the “place theory” of human hearing. The filter outputs of the series of bands correlate to certain phonetically important speech sounds. Two types of filter banks popularly used are critical band digital filter banks and fourier transform derived filter banks [14] amplitudes. Both methods transform the linear acoustic frequency into perceptually motivated frequency scales such as bark scale or mel scale that are formulated based on phycoacoustical and hearing study. Mel-scale is more popular approximation to the mechanism of how basilar membrane decomposes pure tones in logarithmic fashion.

In the past, researchers used arbitrary mel-frequency resolutions such as using 20 filters [21] and 40 filters [22] which turned out to give different mel resolution. On the other hand, the aim of this paper is to investigate which critical bands of the fourier transform derived mel-filters contain phonetically important ethnical information of the MalE database. This approach proposes an innovative way of selecting salient features by implementing a simple selection algorithm to select bands based on smallest variances within class scores. It is hoped that this work can optimize both the presentation of speech features and the performance of accent classification system.

This paper is organized as follows. In section II, we describe about experimental setup and speech database. Section III explains about extraction of speech features, feature selection using the proposed statistical band selection approach and accent modeling. Results of the conducted experiments are discussed in section IV. Lastly, section V concludes the important findings of this paper.

II. Speech Corpus Database and Experimental Setup

The experiments were conducted on the local MalE database consists of selected accent-sensitive wordlist in our previous analysis [23]. For the analysis purpose of this work, we took speech corpus recorded from 45 female volunteers and 45 male volunteers of three main ethnics. It was composed of 15 Malay, 15 Chinese and 15 Indian speakers of each gender. The accent sensitive words are *bottom*, *aluminum* and *target* and each word was replicated five times for each speaker. This collection of utterances amounted to 1350 speech samples. The speakers were originated from various north, south, west and east regions of the country and as such they were also influenced by their regional accents. Subjects were diploma, undergraduate and postgraduate students of Universiti Malaysia Perlis aged from 18 to less than 35 years. The recording was carried in a semi-anechoic acoustic chamber as shown in Fig. 1 having background noise of approximately 22dB. The recordings were held using a condenser, supercardioid and unidirectional microphone and a laptop computer sound card with MATLAB program. The sampling rate and bit resolution were set to 16kHz and 16bps respectively. Table 1 summarizes the description of the MalE database.

TABLE I
DATABASE DESCRIPTION

Accent	Gender	No of Speakers	No of utterances (N)
Malay	Male	15	225
	Female	15	225
	Total	30	450
Chinese	Male	15	225
	Female	15	225
	Total	30	450
Indian	Male	15	225
	Female	15	225
	Total	30	450
Total	Male	45	675
	Female	45	675
	Total	90	1350

Fig. 1. Experimental setup for speech recording in semi-anechoic chamber.

III. Methodology

This section explains about the steps and algorithms for extracting features, the proposed statistical band selection and accent classification model used in this work.

A. Mel-filter bands Spectral Energy

The block diagram describing the procedures of mel-bands spectral energy extraction is depicted in Fig. 2. The speech data was initially zero-adjusted to remove a DC bias during recording and pre-emphasized using first-order FIR filter with transfer function of $H(z)=1-0.9375z^{-1}$ to compensate the attenuation in spectral energy approximately by 6dB per octave. Next, the speech was frame-blocked into 32msec short-time frames with 50% overlapping and Hamming windows were then applied to the frames. The normal frame length for male and female voices is between 20msec to 40msec to ensure stationary property [24]. The Hamming window function is expressed mathematically in (1).

$$w(n)=0.54-0.46\cos\left[2\pi n/(N-1)\right]. \quad (1)$$

where $0 \le n \le N-1$ is the sample point in the short-time windowed frames and N is the window length which is equal to the frame length.

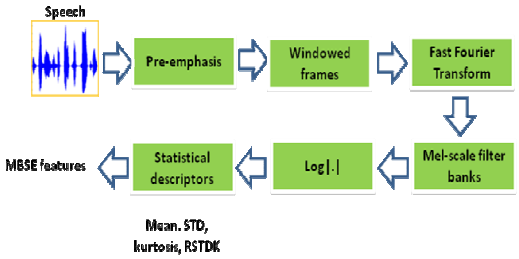

Fig. 2. Block diagram of mel-bands spectral energy feature extraction using statistical descriptors.

Next, the speech spectrum was computed from the pre-processed frames using the fast fourier transform (FFT) algorithm. Considering a better way than linear frequency scale as proposed by human perceptual study, this frequency range was warped using a new scale known as mel-scale, the known variation of the human ear's critical bandwidths with frequency [14]. In this scale, filters are uniformly spaced on the perceptually motivated scale while nonlinearly spaced in the linear scale. The bandwidths overlap with each other by 50% as can be seen in Fig. 3. The filters are densely located for low frequencies but sparsely located for high frequencies to emphasize the lower frequency components are more important in speech analysis. The transformation of the center frequencies in Hz from the linear to the mel-scale is expressed in (2).

$$m_f = 2595 \log_{10}\left(1 + f/700\right). \quad (2)$$

where m_f denotes the resulted frequency in the mel scale and f denotes the corresponding frequency in the linear scale.

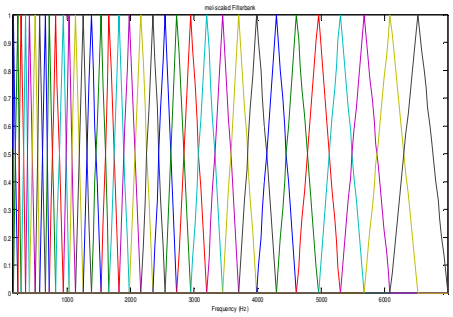

Fig. 3. Mel filter banks basis functions.

Applying mel-filter banks in frequency domain is simply amounts to multiplying triangular-shaped windows to each region of spectrum of interest. The spectral energy of 18 mel-bands were computed taking four statistical measures i.e. mean, standard deviation (STD), kurtosis and the ratio of STD to kurtosis (RSTDK) of each word sample to describe the distribution of our data [25]. The log-energy output of each mel-band is calculated as in (3). The resulted acoustic feature vector is onwards spelt as mel-bands spectral energy (MBSE).

$$W(i) = \sum \log|E_k| . H_i\left(\frac{2\pi k}{N}\right). \quad (3)$$

where $W(.)$ denotes the output of mel-warped log energy for the i^{th} critical band filter, $E(.)$ refers to the FFT power spectrum and $H_i(.)$ is the transfer function of the i^{th} mel-filter with k as the FFT sample index and N is the number of filters in the filter banks. This summation is done between the lower and upper frequencies of each filter with nonzero coefficients.

B. Mel-frequency Cepstral Coefficients

A block diagram showing the steps involve in extracting mel-frequency cepstral coefficients (MFCC) is depicted in Fig. 4 follows the same steps as MBSE except for the last stage. The discrete fourier transform (DCT) replaced the statistical descriptors to derive cepstral that is defined as local spectrum property of the speech signal. Cepstrum is a new domain called as quefrency as the frequency transform was applied for the second time to convert back to the time-like domain after liftering.

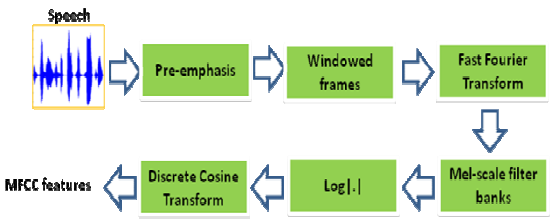

Fig. 4. Block diagram of MFCC feature extraction.

The cepstral coefficients of mel-scale filter banks [26] can be computed as in (4) by summing all the product of fourier transform derived log-

energy output of individual bandpass filter and DCT.

$$C_m = \sum_{k=1}^{N} E_k \cos\left[m(k-0.5)\pi / N\right]. \quad (4)$$

where variables *C(.)* and *E(.)* represent the m^{th} cepstral coefficient (cepstrum) and k^{th} log-energy respectively. *N* is the number of filters in the filter banks and the number of cepstrum takes in this order i.e. *m=1, 2, …, M*.

The lower order cepstrum represents the slowly varying part of the spectrum while spikes in the series correspond to the harmonic series of the vocal folds [27]. Normally a few lower order coefficients are taken to represent the vocal tract shape normally *M*=12 or 13 is sufficient to leave out the pitch property of the speech signal.

C. Statistical Band Selection

The easiest way to select generated statistical MBSE features from a number of the filter bank outputs is to measure the spread of scores in each band of the mel-filter banks. The variances were measured as the objective measures to be sorted in ascending order. A simple algorithm named as statistical band selection (SBS) to select the mel-bands is described as follow:

Step 1: For each accent group, extract only the mean scores of mel-bands log energies in the feature vectors MBSE=$[f_{11}, f_{12}, f_{13}, f_{14}, f_{21}, f_{22}, f_{23}, f_{24}, \ldots, f_{ij}, \ldots, f_{IJ}]^T$ where *i* and *j* represent the index for band number (1, 2, 3, … , *I*) and statistical feature types (1=mean, 2=STD, 3=kurtosis), 4=RSTK) with the maximum parameters as *I*=18 and *J*=4. Only the first feature (f_{i1}) of each band will be extracted for the objective measure of the band selection.

Step 2: For each band, calculate the variance of the mean scores of all word samples as in (5).

$$V(f_{i1}) = \frac{1}{N}\sum_{k=1}^{N}\left(f_{i1k} - \overline{f_i}\right)^2. \quad (5)$$

where the variables *V(.), f(.)* and $\bar{f}(.)$ represent the band variance of the log-energy mean scores, word sample mean score and the mean value of the mean scores of the i^{th} band respectively. The bands are numbered in ascending order as $i = \{1, 2, 3, ..., I\}$ and *k*=1, 2, 3, …, *N* where *N* is the word sample size.

Step 3: Sort the variances in ascending order with the smallest value appears in the first column followed by the second smallest next to it and so forth expressed mathematically as in (6).

$$V\left(f_{i'1_1}\right) \le V\left(f_{i'1_2}\right) \le \ldots\ V\left(f_{i'1_I}\right). \quad (6)$$

Next, re-map the original *i* to a new band order i' based on the index appeared in (7). The feature vector arrangement now will follow this new order as in (7).

$$i'_{Gl} = \{i'_{Gl1}, i'_{Gl2}, i'_{Gl3}, \ldots, i'_{Gll}\}. \quad (7)$$

where *l* is the group label and defined as $l = \{1, 2, 3\} = \{\text{'Malay', 'Chinese', 'Indian'}\}$.

Step 4: Repeat Step 1 to 3 for the other two accent groups. Save the resulted new band order as i'_{G1}, i'_{G2} and i'_{G3} for three different accent groups respectively.

Step 5: Select the band features that are common in the three accent groups i.e. $i'_{G1} \cap i'_{G2} \cap i'_{G3}$ and formulate the feature set.

Step 6: Model the artificial neural network based on the reduced feature vector length based on the resulted feature set in *Step 5*.

D. Speaker Accent Classifier

In this work we used artificial neural network (ANN) to perform the nonlinear task of the speech signal. Error Back-propagation algorithm is the most popular one used to train Feed-forward Multilayer Perceptron (FF-MLP) for most classification problems [28]. Fig. 5 shows the architecture of the FF-MLP used in this paper. We used two-layer FF-MLP to classify the input features into one of three accent classes and the network was trained using Levenberg-Marquardt learning algorithm that is known for its fast convergence. We adopted mean-squared error (MSE) as an objective criterion for successful learning of the task. The training performance based on MSE can be viewed as in Fig. 6. The hidden and output neurons were both using binary sigmoidal or logistic function. The mathematical expression for the activation function is expressed in (8). The input was normalized between 0.1 to 0.9 using expression in (9) to assist the learning process.

$$f(net) = 1/\left(1 + e^{-net}\right). \quad (8)$$

where *net* is the weighted sum of input to a neuron and *f(net)* is the output.

$$x' = 0.8\left(x - x_{\min}\right)/\left(x_{\max} - x_{\min}\right) + 0.1. \quad (9)$$

where x is the input to input neurons, x_{min} is the minimum value and x_{max} is the maximum value of the feature set and x' is the normalized input.

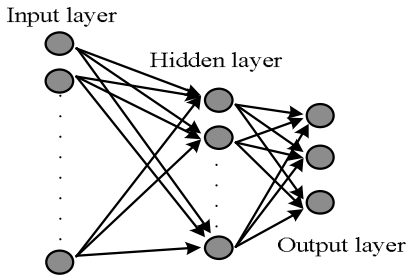

Fig. 5. The architecture of two-layer feed-forward multilayer perceptron (FF-MLP) neural network.

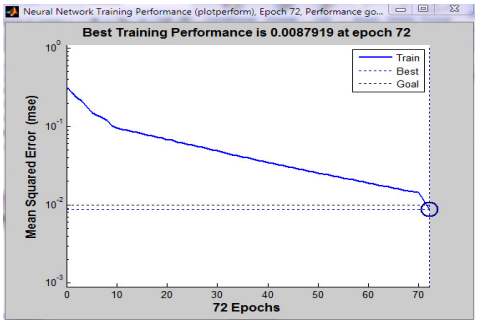

Fig. 6. Training performance of neural network using Mean-squared Error (MSE) as the objective of learning task.

IV. Results and Discussion

This section analyzes the proposed SBS based on the simple algorithm given in the methodology section. ANN accent classifier was adopted as the engine to classify the baseline full-bands MBSE (so forth referred to as BS-MBSE) and the selected bands MBSE (so forth referred to as SBS-MBSE) input features into one of the three accent classes i.e. the Malay accent, the Chinese accent and the Indian accent. The performance was measured using percentage of classification rates (CRs) for individuals and class-overall CRs, feature size, and ANN training time in seconds.

In the first experiment, the band selections for both genders based on the SBS algorithm were employed and analyzed. Ensuing this, BS-MBSE and SBS-MBSE features were used to model the ANN accent classifiers in the second experiment. In the third experiment, MFCC-based features were derived from the full bands and selected bands to explore for further effects of the proposed SBS method. Lastly, in the fourth experiment, we repeated the tests using statistical *K*-Nearest Neighbors (KNN) as alternative classifier.

Different analyses were performed for male and female speakers separately assuming that gender classification has been performed prior to accent classification. This is important to isolate the problem under study since speech is overriding by many attributes simultaneously.

A. Experiment 1

The results of band ranking based on the SBS algorithm of the total 18 mel-bands are tabulated in Table IV and V in ascending order of variances for both genders. After selecting the bands in *Step 5* of the algorithm, we re-mapped the band number to i^{th} index and achieved two stages of band selection named as SBS-MBSE-44 and SBS-MBSE-56 as shown in Table II for the female speakers. Table III complementarily shows the respective results on the male speakers with the resulted SBS-MBSE-32 and SBS-MBSE-40.

TABLE II
STATISTICAL BAND SELECTION RESULTS OF FEMALE DATASET

Bands selection method	Feature vector length	Band no. (*i*) in the set	Band no. (*i*) omitted
Stage 1: SBS-MBSE-44	11x4=44	{6, 7, 8, 9, 10, 13, 14, 15, 16, 17, 18}	{1, 2, 3, 4, 5, 11, 12}
Stage 2: SBS-MBSE-56	14x4=56	{5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18}	{1, 2, 3, 4}

TABLE III
STATISTICAL BAND SELECTION RESULTS OF MALE DATASET

Bands selection method	Feature vector length	Band no. (*i*) in the set	Band no. (*i*) omitted
Stage 1: SBS-MBSE-32	8x4=32	{6, 7, 8, 13, 15, 16, 17, 18}	{1, 2, 3, 4, 5, 9, 10, 11, 12, 14}
Stage 2: SBS-MBSE-40	10x4=40	{6, 7, 8, 9, 13, 14, 15, 16, 17, 18}	{1, 2, 3, 4, 5, 10, 11, 12}

For the female speakers, the first stage band selection i.e. SBS-MBSE-44 resulted 11 bands in common for the three datasets (Malay, Chinese and Indian accents). The results showed that a more important frequency contents that best characterized each dataset ranged from 921.5 Hz to 2254.5 Hz and 3055.9 Hz to 7016.2 Hz. This resulted in a dimension size of 44 of MBSE feature vector. A further search for common bands resulted another 3 bands totaling to 14 bands selected in the second stage i.e. SBS-MBSE-56. This embodied extended frequency contents of 738.1 Hz to 7016.2 Hz. We concluded from this experiment that the lower part of frequency bands from 189.9 Hz to 738.1 Hz were less important that can be neglected in the input features of the female speakers.

For the male speakers, the first stage band selection i.e. SBS-MBSE-32 resulted 8 bands in common for the three datasets (Malay, Chinese and Indian accents). Nevertheless, in the first stage selection of 8 common bands, the 13^{th}-band was located at the 12^{th} sorting position for the Indian male dataset, which was outside the realm of the first eight selected bands in the first stage selection. Thus, the chosen important frequency contents that best characterized the three datasets were in the range of 921.5 Hz to 1624.1 Hz, 3055.9 Hz to 3534.7 Hz and 4074.7 Hz to 7016.2 Hz. This resulted in a dimension size of 32 of MBSE feature vector. A further search for common bands resulted another 2 bands totaling to 10 bands selected in the second stage i.e. SBS-MBSE-40. This resulted in extended frequency contents of 921.5 Hz to 1920.4 Hz and 3055.9 Hz to 7016.2 Hz. We concluded from this experiment that the lower part of frequency bands from 189.9 Hz to 921.5 Hz and 1920.4 Hz to 3055.9 Hz were less important that can be neglected in the input features of the male speakers.

B. Experiment 2

In the next experiment, we modeled ANN with n=72, 44 and 56 input neurons for the BS-MBSE-72, SBS-MBSE-44 and SBS-MBSE-56 feature sets for the female dataset whilst the BS-MBSE-72, SBS-MBSE-32 and SBS-MBSE-40 for the male dataset with n=72, 32 and 40 respectively. The output layer has m=3 nodes wherein activation of a node represents an accent individually and the number of hidden neurons p=35 was fixed by regress analysis on that parameter.

To cater for random fluctuations of neural network's weights nature, we trained 10 independent networks using 80% of total samples i.e. 540 samples in the training phase with different sets of initial weights. Consequently, using the remaining 20% i.e. 135 samples in each trial as test dataset, we obtained CRs to evaluate the performance of the networks. In the testing phase, we adopted threshold and margin criterion as suggested by [29] to declare the state of output neurons to be one for simulator's result of above 0.9 and below 0.1 to be zero. With this in mind we set testing tolerance of 0.1 in both limits. Output values between these limits were considered marginal and labeled as unclassified or inconclusive results. These were not counted in the classification rate. The learning rate (α) and momentum rate (β) were fixed to 0.5 and 0.9 respectively to accelerate the training convergence and also to avoid local minimum trap via trial and error for all feature sets.

Table VI–VIII show the performance of ANN accent classifier with different input features of the BS-MBSE and SBS-MBSE for both female and male speakers. Apart from the classification rates, the training time taken for each feature type was recorded to measure the speed of those input-based systems.

The minimum (min), maximum (max), mean and standard deviation (STD) values of CRs and training time were computed for 10 trials of the neural network models and recorded in the bottom rows of the aforementioned tables. From Table VI, it is observed that in average the male speakers gained better accuracy than the female speakers by 4.72% with 30.84% lesser deviation from the mean value for the whole trials. Nevertheless, it required more time of 38.82% longer to train the network in average but the difference was about 57 sec. The max and min CRs recorded for the males were higher i.e. 99.08% and 94.12% as compared to the females of 95.70% and 89.22% respectively. The accuracy for Indian males was higher than the others and both Chinese males and females were recognized less accurately than others by drops of approximately 2% to 5% and 7% respectively.

The efficacy of our proposed SBS method in selecting certain bands to improve the performance of the system has been achieved by

an increase of the overall CR of 1.85% using SBS-MBSE-56 for the female dataset with unchanged training speed of 145 sec. In terms of feature size, it has been reduced by 22.22%. The experiment has achieved similar individual class CR of about 94%. For the male dataset, SBS-MBSE-32 has proven to maintain the CR of approximately 96% but with 70.60% faster in speed. The feature size also has been reduced by 55.56% using this method as compared to using the full-band BS-MBSE. The Malay and Indian classes yielded CRs of about 97% each as compared to the Chinese class, yielded CR of about 94%. All results are shown in Table VII and VIII.

Fig. 7 and 8 visually compare the performance of BS-MBSE-72 (full-bands), handcrafted reduced bands BS-MBSE and the band selection using SBS method in terms of the average of overall CRs since this measure is more important relating to the performance.

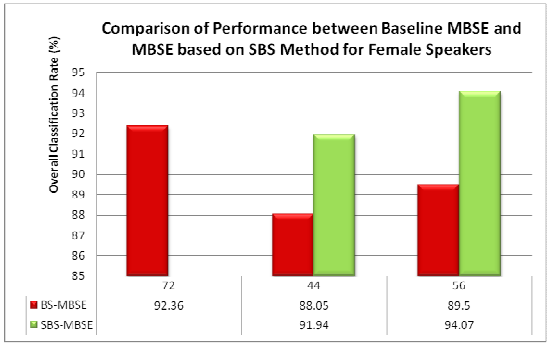

Fig. 7. Performance of the baseline features (BS-MBSE) and the statistical band selection methods i.e. SBS-MBSE-44 and SBS-MBSE-56 having different size of feature vector using ANN for the female dataset.

To compare the performance of ANN if no selection had been applied, the BS-MBSE was simply taken as the first 56 features and the first 44 features in the BS-MBSE with no selection method for the females and similarly, taken the first 40 features and the first 32 features for the males. Obviously with proper method of band selection, SBS-MBSE outperformed BS-MBSE by approximately 4% to 6% for both genders. All results displayed were the overall CRs averaged over 10 trials.

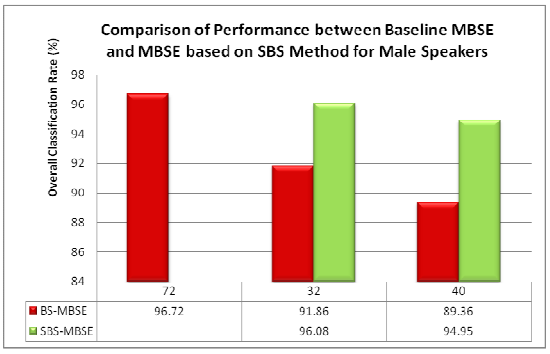

Fig. 8. Performance of the baseline features (BS-MBSE) and the statistical band selection methods i.e. SBS-MBSE-32 and SBS-MBSE-40 having different size of feature vector using ANN for the male dataset.

C. *Experiment 3*

Further effects on the propose SBS method were tested on MFCC features by taking cepstrals using DCT instead of using statistical descriptors as in the previous work. Table IX–XI show the complete results for MFCC derived from the baseline full bands (so forth referred to as BS-MFCC13) and the selected bands (so forth referred to as SBS-MFCC13). Number of coefficients used was 13 with reference to many past researches. The best accuracy achieved using MFCC derived from full-bands utilization for the females and males were 94.06% and 98.15% while the average CRs were 89.07% and 95.69% respectively. These records were a bit lower than those obtained using the descriptors of MBSE for both cases of gender. However the speed has been increased by as much as 84.71% to 97.38% with a reduction in the feature size from 72 to 13 for females and males using full bands features. By applying SBS on band selection, new feature sets were derived i.e. SBS-MFCC13-44 and SBS-MFCC13-56 for the females and SBS-MFCC13-32 and SBS-MFCC13-40 for the males. Among these, SBS-MFCC13-56 has proven useful by increasing the CR by 3.58% and faster training speed of 70.85% as compared to BS-MFCC13-72.

Fig. 9 shows all the performance using the baseline full-bands and the selected bands based on SBS method in terms of the average of overall CRs for each gender. Only the best SBS-MBSE and SBS-MFCC13 are shown for comparison purpose. The selected feature sets were SBS-MBSE-32 and SBS-MFCC13-40 for the males while that of SBS-MBSE-56 and SBS-MFCC13-

56 for the females. Overall, the male speakers outperformed the female speakers on all feature sets except for the selected bands MFCC13. It seemed that the males did not take much advantage in terms of CRs using SBS method except for speed and feature size. Secondly, descriptors of MBSE classified better than MFCC but at greater cost of memory requirement and slower. Thirdly, the proposed SBS method has proven useful in improving the performance of both descriptors of MBSE- and MFCC-based ANN classifiers.

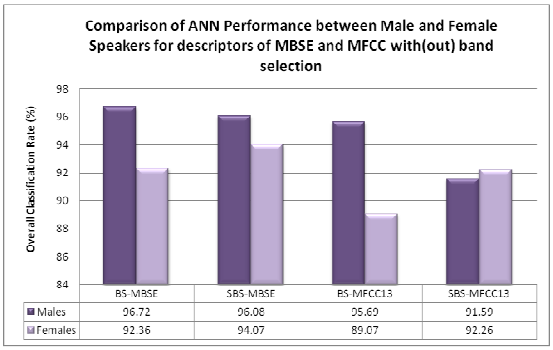

Fig. 9. Performance of the male and female speakers using baseline features and statistical band selected features on descriptors of MBSE and MFCC using ANN.

D. Experiment 4

The experiments were repeated using KNN with K=2 recorded as the best results shown in Fig. 10. The results owned consistent explanation as using ANN previously. In all cases, it could be seen that SBS-MBSE was more efficient than BS-MBSE in both male and female datasets.

Lastly we compared the performance of ANN and KNN for using BS-MBSE, SBS-MBSE, BS-MFCC13 and SBS-MFCC13 based classifiers. Fig. 11 and 12 show the results for male dataset and female dataset respectively. Obviously, ANN outperformed KNN in almost all cases. Again, only the best SBS-MBSE based descriptors and MFCC are shown for comparison purpose as described previously.

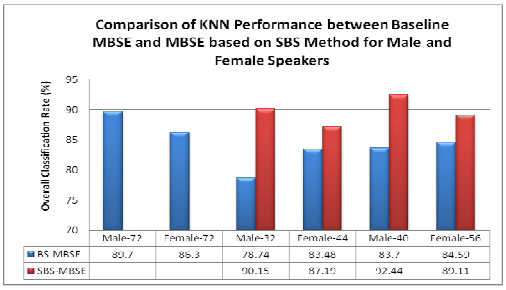

Fig. 10. Performance of the baseline features (BS-MBSE) and the statistical band selection (SBS-MBSE) methods using KNN for males and female speakers.

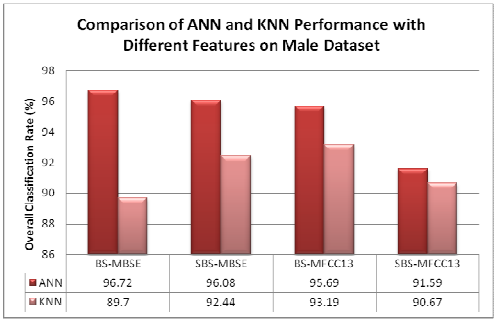

Fig. 11. Performance of the male speakers using baseline features and statistical band selected features on descriptors of MBSE and MFCC using ANN and KNN.

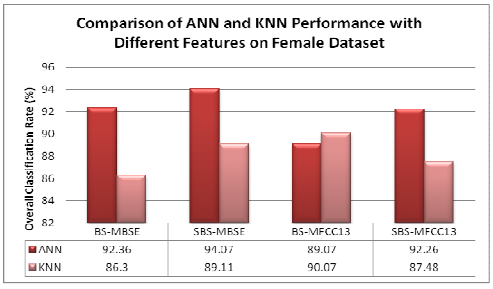

Fig. 12. Performance of the female speakers using baseline features and statistical band selected features on descriptors of MBSE and MFCC using ANN and KNN.

V. CONCLUSION

This paper has presented three-class accent classification in Malaysian English accented speech utilizing features from optimized mel-bands generated using the proposed statistical band selection (SBS) algorithm. This algorithm sorted the variances of within class mean scores in ascending order to choose common bands in all accent groups. Two types of features employed from the baseline full-bands and the selected

bands namely descriptors of mel-bands spectral energy (MBSE) and mel-frequency cepstral coefficients (MFCC). Two baseline features were BS-MBSE-72 and BS-MFCC13-72 while eight new generated features from SBS algorithm were SBS-MBSE-32, SBS-MBSE-40, SBS-MFCC13-32 and SBS-MFCC13-40 for the male dataset while SBS-MBSE-44, SBS-MBSE-56, SBS-MFCC13-44 and SBS-MFCC13-56 for the female dataset.

Experimental results have proven the efficacy of the proposed SBS method on MBSE features over no selection method and the baseline BS-MBSE by approximately 4% to 6% for both genders and reduction in feature size of 22% to 55%. Secondly, descriptors of MBSE classified better than MFCC but at greater cost of memory requirement and slower by approximately 84.71% to 97.38%. Thirdly, either using ANN or KNN, the proposed SBS method has proven useful in improving the performance of both descriptors of MBSE- and MFCC-based ANN classifier in terms of classification rate, memory space and speed. Lastly, it was found that ANN outperformed KNN in most cases.

ACKNOWLEDGMENT

The authors would like to acknowledge the support and encouragement given by the Vice Chancellor of University Malaysia Perlis (UniMAP) that is Brig. Jeneral Dato' Prof. Dr. Kamaruddin Hussin and the sponsorship in Ph.D candidature from the Ministry of Higher Education, Malaysia.

TABLE IV
RE-SORTING THE BAND NUMBER USING STATISTICAL BAND SELECTION ALGORITHM FOR FEMALE MALAYSIAN ENGLISH DATASETS

Bank ranking according to variances (the smallest on the top)								
Malay			**Chinese**			**Indian**		
Band no. (*i*)	**Variance**	**Frequency contents (Hz)**	**Band no. (*i*)**	**Variance**	**Frequency contents (Hz)**	**Band no. (*i*)**	**Variance**	**Frequency contents (Hz)**
18	0.2739	6143.7 – 7016.2	15	0.3603	4074.7 – 4683.4	18	0.3718	6143.7 – 7016.2
16	0.2815	4683.4 – 5369.8	16	0.4235	4683.4 – 5369.8	17	0.4387	5369.8 – 6143.7
17	0.2896	5369.8 – 6143.7	14	0.4588	3534.7 – 4074.7	8	0.5022	1361.3 – 1624.1
15	0.3018	4074.7 – 4683.4	8	0.4871	1361.3 – 1624.1	9	0.5113	1624.1 – 1920.4
8	0.3814	1361.3 – 1624.1	17	0.5374	5369.8 – 6143.7	15	0.5271	4074.7 – 683.4
9	0.4126	1624.1 – 1920.4	7	0.5471	1128.2 – 1361.3	7	0.5534	1128.2 – 1361.3
14	0.4613	3534.7 – 4074.7	18	0.5563	6143.7 – 7016.2	14	0.5683	3534.7 – 4074.7
7	0.4745	1128.2 – 1361.3	9	0.5849	1624.1 – 1920.4	16	0.5854	4683.4 – 5369.8
10	0.4962	1920.4 – 2254.5	13	0.5946	3055.9– 3534.7	10	0.6265	1920.4 – 2254.5
13	0.5596	3055.9– 3534.7	6	0.6701	921.5 – 1128.2	6	0.7102	921.5 – 1128.2
6	0.5774	921.5 – 1128.2	10	0.6916	1920.4 – 2254.5	13	0.7284	3055.9– 3534.7
12	0.6027	2631.2 – 3055.9	12	0.7123	2631.2 – 3055.9	11	0.8053	2254.5 – 2631.2
11	0.6333	2254.5 – 2631.2	5	0.7427	738.1 – 921.5	12	0.8091	2631.2 – 3055.9
5	0.7036	738.1 – 921.5	11	0.7596	2254.5 – 2631.2	5	0.8807	738.1 – 921.5
3	0.7193	431.3 – 575.5	2	0.7944	303.3 – 431.3	3	0.8993	431.3 – 575.5
4	0.8227	575.5 – 738.1	1	0.8183	189.9 – 303.3	2	0.9817	303.3 – 431.3
1	0.9129	189.9 – 303.3	4	0.9308	575.5 – 738.1	4	1.0291	575.5 – 738.1
2	0.9867	303.3 – 431.3	3	0.9430	431.3 – 575.5	1	1.0585	189.9 – 303.3

TABLE V
RE-SORTING THE BAND NUMBER USING STATISTICAL BAND SELECTION ALGORITHM FOR MALE MALAYSIAN ENGLISH DATASETS

Bank ranking according to variances (the smallest on the top)								
Malay			**Chinese**			**Indian**		
Band no. (*i*)	**Variance**	**Frequency contents (Hz)**	**Band no. (*i*)**	**Variance**	**Frequency contents (Hz)**	**Band no. (*i*)**	**Variance**	**Frequency contents (Hz)**
18	0.2289	6143.7 – 7016.2	18	0.2941	6143.7 – 7016.2	16	0.2963	4683.4 – 5369.8
17	0.2477	5369.8 – 6143.7	17	0.3417	5369.8 – 6143.7	17	0.3650	5369.8 – 6143.7
16	0.3274	4683.4 – 5369.8	16	0.4507	4683.4 – 5369.8	18	0.4186	6143.7 – 7016.2
7	0.4032	1128.2 – 1361.3	7	0.4795	1128.2 – 1361.3	15	0.4188	4074.7 – 4683.4
8	0.4812	1361.3 – 1624.1	15	0.4844	4074.7 – 4683.4	6	0.4397	921.5 – 1128.2
13	0.4901	3055.9– 3534.7	8	0.5115	1361.3 – 1624.1	7	0.4602	1128.2 – 1361.3
6	0.4969	921.5 – 1128.2	13	0.5129	3055.9– 3534.7	8	0.5407	1361.3 – 1624.1
15	0.5205	4074.7 – 4683.4	6	0.5520	921.5 – 1128.2	5	0.5690	738.1 – 921.5
14	0.5504	3534.7 – 4074.7	14	0.5562	3534.7 – 4074.7	14	0.5770	3534.7 – 4074.7
12	0.5996	2631.2 – 3055.9	12	0.5854	2631.2 – 3055.9	4	0.6162	575.5 – 738.1
1	0.6580	189.9 – 303.3	9	0.6390	1624.1 – 1920.4	3	0.625	431.3 – 575.5
9	0.6609	1624.1 – 1920.4	11	0.6610	2254.5 – 2631.2	13	0.6613	3055.9– 3534.7

3	0.7541	431.3 – 575.5	5	0.6666	738.1 – 921.5	1	0.6663	189.9 – 303.3
2	0.7670	303.3 – 431.3	10	0.7107	1920.4 – 2254.5	9	0.6675	1624.1 – 1920.4
11	0.7693	2254.5 – 2631.2	4	0.7812	575.5 – 738.1	2	0.6692	303.3 – 431.3
5	0.7792	738.1 – 921.5	1	0.8233	189.9 – 303.3	12	0.6741	2631.2 – 3055.9
4	0.8121	575.5 – 738.1	3	0.8678	431.3 – 575.5	11	0.7198	2254.5 – 2631.2
10	0.8396	1920.4 – 2254.5	2	0.8827	303.3 – 431.3	10	0.7597	1920.4 – 2254.5

TABLE VI
PERFORMANCE OF ANN ACCENT CLASSIFIER USING BASELINE MEL-BAND SPECTRAL ENERGY (BS-MBSE-72)

Gender	**Female**					**Male**				
Trial	**Classification rate (%)**				**Training time (s)**	**Classification rate (%)**				**Training time (s)**
	Malay	Chinese	Indian	Overall		Malay	Chinese	Indian	Overall	
1	90.91	83.78	94.59	89.72	55	94.12	93.94	100	96.15	135
2	96.67	85.71	90.70	91.09	67	93.55	90.63	97.44	94.12	210
3	96.55	90.63	94.74	93.94	117	93.75	90.91	97.73	94.5	166
4	88.89	93.33	97.50	93.81	283	100	91.43	100	97.12	161
5	93.33	80.56	95.00	89.62	56	100	93.75	100	98.04	159
6	93.10	87.88	97.22	92.86	245	96.77	93.75	97.22	95.96	173
7	96.00	90.32	95.12	93.81	133	93.55	96.77	100	97.12	197
8	100	90.32	96.88	95.70	136	100	97.22	100	99.08	353
9	93.75	84.38	89.47	89.22	256	100	96.77	97.50	98.02	212
10	92.31	91.43	97.22	93.81	110	93.55	96.97	100	97.12	258
Min	88.89	80.56	89.47	**89.22**	55	93.55	90.63	97.22	**94.12**	135
Max	100	93.33	97.50	**95.70**	285	100	97.22	100	**99.08**	353
Mean	94.15	87.83	94.84	**92.36**	**145.8**	96.53	94.21	98.99	**96.72**	**202.4**
STD	3.22	4.07	2.76	**2.27**	85.4	3.13	2.62	1.31	**1.57**	63.5

TABLE VII
PERFORMANCE OF ANN ACCENT CLASSIFIER USING STATISTICAL BAND SELECTION (SBS-MBSE) – FIRST STAGE

Gender	**Female (SBS-MBSE-44)**					**Male (SBS-MBSE-32)**				
Trial	**Classification rate (%)**				**Training time (s)**	**Classification rate (%)**				**Training time (s)**
	Malay	Chinese	Indian	Overall		Malay	Chinese	Indian	Overall	
1	93.33	96.55	93.61	94.34	25	100	96.97	100	98.98	73
2	95.83	96.67	97.44	96.77	50	96.77	93.75	97.22	95.96	71
3	93.55	87.10	97.62	93.27	115	97.37	91.18	96.97	95.24	49
4	95.24	93.75	91.43	93.18	72	94.44	89.74	100	94.69	43
5	92.59	93.94	95.24	94.12	90	100	91.89	96.88	96.11	67
6	91.67	79.41	91.89	87.85	97	96.67	93.94	93.94	94.79	49
7	96.97	90.91	94.12	94.00	149	90.91	96.97	94.59	94.17	37
8	82.86	90.91	94.74	89.62	176	96.88	100	97.22	98.00	128
9	83.33	82.35	80.95	82.14	47	100	91.67	100	96.91	55
10	100	83.87	97.37	94.12	189	100	93.75	94.44	95.96	23
Min	82.86	79.41	80.95	82.14	25	90.91	89.74	93.94	94.17	23
Max	100	96.67	97.62	96.77	189	100	100	100	98.98	128
Mean	92.54	89.55	93.44	91.94	101	97.30	93.99	97.13	**96.08**	**59.5**
STD	5.52	6.08	4.91	4.27	55.9	2.96	3.16	2.33	**1.52**	28.7

TABLE VIII
PERFORMANCE OF ANN ACCENT CLASSIFIER USING STATISTICAL BAND SELECTION (SBS-MBSE) – SECOND STAGE

Gender	**Female (SBS-MBSE-56)**					**Male (SBS-MBSE-40)**				
Trial	**Classification rate (%)**				**Training time (s)**	**Classification rate (%)**				**Training time (s)**
	Malay	Chinese	Indian	Overall		Malay	Chinese	Indian	Overall	
1	93.10	90.32	91.89	91.75	111	97.30	90.91	96.67	95.00	72
2	100	96.30	94.29	96.81	154	93.10	89.19	100	94.06	80
3	91.43	88.24	94.12	91.26	84	96.97	92.31	93.75	94.23	43
4	93.75	96.15	97.14	95.70	204	89.66	91.89	100	94.33	31
5	86.11	97.06	91.89	91.59	126	100	89.19	100	96.15	15
6	100	90.32	92.11	94.11	103	100	89.07	96.43	94.79	91
7	90	100	97.06	95.74	163	96.77	91.43	97.22	95.10	47
8	100	96.30	97.06	97.85	225	96.97	90	96.88	94.29	69
9	92.31	97.06	89.74	92.86	150	94.12	94.44	100	96.30	112
10	100	92.86	94.12	95.74	139	97.44	87.88	100	95.28	67
Min	86.11	88.24	89.74	91.26	84	89.66	87.88	93.75	94.06	15
Max	100	100	97.14	97.85	225	100	94.44	100	96.30	112
Mean	94.67	94.46	93.94	**94.07**	**145.9**	96.23	90.63	98.10	94.95	62.7
STD	5.03	3.79	2.57	**2.33**	43.9	3.16	1.95	2.21	0.79	29.0

TABLE IX
PERFORMANCE OF ANN ACCENT CLASSIFIER USING MFCC OF BASELINE FULL BANDS (BS-MFCC13-72)

Gender	**Female**					**Male**				
Trial	**Classification rate (%)**				**Training time (s)**	**Classification rate (%)**				**Training time (s)**
	Malay	Chinese	Indian	Overall		Malay	Chinese	Indian	Overall	
1	100	87.1	91.67	93.52	5	93.33	95.35	92.31	93.75	4
2	88.89	78.57	100	89.90	7	100	92.11	100	96.94	6
3	82.86	88.24	100	90.20	148	93.10	92.68	94.59	93.46	7
4	92.86	73.81	83.33	83.33	26	96.43	88.37	94.29	92.45	10
5	95.24	79.49	94.44	89.74	13	100	96.77	97.44	97.96	4
6	77.5	85.71	81.82	81.48	3	100	90.91	97.37	95.74	4
7	92.31	78.38	90.91	87.16	6	100	88.64	100	95.24	7
8	97.44	80	96.43	91.18	8	96.77	92.11	100	95.96	3
9	97.06	91.43	93.75	94.06	4	96.77	100	97.06	98.15	4
10	94.74	81.58	94.29	90.09	3	100	97.37	94.74	97.22	4
Min	77.50	73.81	81.82	81.48	3	93.10	88.37	92.31	92.45	3
Max	100	91.43	100	**94.06**	148	100	100	100	**98.15**	10
Mean	91.89	82.43	92.66	**89.07**	**22.3**	97.64	93.43	96.78	**95.69**	**5.3**
STD	7.00	5.46	6.12	**4.04**	44.7	2.79	3.84	2.73	**1.96**	2.2

TABLE X
PERFORMANCE OF ANN ACCENT CLASSIFIER USING MFCC OF STATISTICAL BAND SELECTION (SBS) – FIRST STAGE

Gender	Female (SBS-MFCC13-44)					Male (SBS-MFCC13-32)				
Trial	Classification rate (%)				Training time (s)	Classification rate (%)				Training time (s)
	Malay	Chinese	Indian	Overall		Malay	Chinese	Indian	Overall	
1	91.67	89.66	87.5	89.69	6	79.31	87.5	90.24	86.27	7
2	89.66	100	86.84	91.21	8	92.86	85.19	88.10	88.66	7
3	97.37	80.00	87.18	89.22	10	72.41	89.66	91.49	85.71	10
4	87.10	93.55	91.89	90.91	7	82.35	81.25	90.48	85.19	7
5	89.19	67.65	85.71	81.13	11	81.82	85.29	95.45	88.29	6
6	79.41	87.50	84.21	83.65	18	84.45	92.31	86.84	87.63	11
7	85.29	94.12	91.89	90.48	6	85.71	85.71	90.70	87.74	9
8	81.25	93.55	85.71	86.73	6	86.67	75.86	91.49	85.85	24
9	87.50	77.14	86.84	83.81	28	78.79	88.46	87.50	84.85	15
10	94.29	86.67	86.84	89.32	4	85.71	85.19	90.00	87.25	14
Min	79.41	67.65	84.21	81.13	4	72.41	75.86	86.84	84.85	6
Max	97.37	100	91.89	91.21	28	92.86	92.31	95.45	88.66	24
Mean	88.27	86.98	87.46	87.62	10.4	83.00	85.64	90.23	86.74	11
STD	5.50	9.62	2.52	3.57	7.3	5.50	4.56	2.45	1.34	5.5

TABLE XI
PERFORMANCE OF ANN ACCENT CLASSIFIER USING MFCC OF STATISTICAL BAND SELECTION (SBS) – SECOND STAGE

Gender	Female (SBS-MFCC13-56)					Male (SBS-MFCC13-40)				
Trial	Classification rate (%)				Training time (s)	Classification rate (%)				Training time (s)
	Malay	Chinese	Indian	Overall		Malay	Chinese	Indian	Overall	
1	87.88	95.83	92.86	92.66	4	100	95.35	70.37	90.83	5
2	89.19	93.62	100	93.52	6	82.35	95.35	90.63	89.91	6
3	81.82	90.24	89.29	87.25	4	97.22	86.96	86.21	90.09	5
4	97.14	84.44	89.66	89.91	9	100	92.68	93.10	95.15	4
5	92.11	95.00	93.75	93.64	6	97.37	89.74	87.10	91.67	5
6	96.97	88.64	92.86	92.38	5	83.33	90.70	89.66	88.23	12
7	94.59	97.67	92.86	95.37	5	91.43	90.00	88.57	90.00	6
8	81.82	92.86	96.97	90.74	13	100	89.74	96.30	94.74	4
9	91.18	88.64	93.10	90.65	6	97.06	93.33	83.33	91.74	5
10	92.31	100	96.43	96.43	7	95.00	94.59	90.32	93.52	5
Min	81.82	84.44	89.29	87.25	4	82.35	86.96	70.37	88.24	4
Max	97.14	100	100	96.43	13	100	95.35	96.30	95.15	12
Mean	90.50	92.69	93.78	**92.26**	**6.5**	94.38	91.84	87.56	91.59	5.7
STD	5.46	4.73	3.27	**2.71**	2.7	6.62	2.84	7.03	2.25	2.3

REFERENCES

[1] J. C. Well, *Accents of English: An Introduction*. New York: Cambridge University Press, 1982.

[2] A. R. Hassan, "Monthly Statistical Bulletin Malaysia," Department of Statistics Malaysia, February, 2012.

[3] H. S. Phoon, "The Phonological Development of Malaysian English Speaking Chinese Children: A Normative Study," in *Speech Sciences*. vol. Doctor of Philosophy Christchurch, New Zealand: University of Canterbury. Communication Disorders, 2010, p. 293.

[4] K. McGee, "Attitudes towards accents of English at the British Council, Penang: What do the students want?," *Malaysian Journal Of ELT Research (MELTA),* vol. 5, pp. 162-205, 2009.

[5] S. Nair Venugopal, "English, identity and the Malaysian workplace," in *World Englishes*. vol. 19 Oxford, UK: Blackwell Publishers Ltd., 2000, pp. 205-213.

[6] D. C. Hart, "Some English Pronunciation Difficulties in Malaysia," *ELT Journal,* vol. 23, p. 270, 1969.

[7] D. Deterding and G. Poedjosoedarmo, "To what extent can the ethnic group of young Singaporeans be identified from their speech?," in *The English Language in Singapore: Research on Pronunciation*, A. Brown, D. Deterding, and E. L. Lo, Eds. Singapore: Singapore Association for Applied Linguistics, 2000, pp. 1-9.

[8] L. Lim, "Ethnic group differences aligned? Intonation patterns of Chinese, Indian and Malay Singaporean English," in *The English Language in Singapore: Research on Pronunciation*, A. Brown, D. Deterding, and E. L. Lo, Eds. Singapore: Singapore Association for Applied Linguistics, 2000, pp. 10 - 21.

[9] D. Deterding, "Measurements of the /eI/ and/AW/ vowels of young English speakers in Singapore," in *The English Language in Singapore: Research on Pronunciation*, A. Brown, D. Deterding, and L. E. Ling, Eds. Singapore: Singapore Association for Applied Linguistics, 2000, pp. 93–9.

[10] D. Deterding, "The vowels of the different ethnic groups in Singapore," in *English in Southeast Asia Varieties, Literacies and Literatures*, D. Prescott, A. Kirkpatrick, I. Martin, and A. Hashim, Eds. Newcastle UK: Cambridge Scholars Press, 2007, pp. 2-29.

[11] L. M. Arslan and J. H. L. Hansen, "Language accent classification in American English," *Speech Communication,* vol. 18, pp. 353-367, 1996.

[12] J. J. Humphries, P. C. Woodland, and D. Pearce, "Using accent-specific pronunciation modelling for robust speech recognition," in *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, 1996, pp. 2324-2327 vol.4.

[13] S. Dupont, C. Ris, O. Deroo, and S. Poitoux, "Feature extraction and acoustic modeling: an approach for improved generalization across languages and accents," in *Automatic Speech Recognition and Understanding, 2005 IEEE Workshop on*, 2005, pp. 29-34.

[14] J. W. Picone, "Signal modeling techniques in speech recognition," *Proceedings of the IEEE,* vol. 81, pp. 1215-1247, 1993.

[15] J. W. Pitton, W. Kuansan, and J. Biing-Hwang, "Time-frequency analysis and auditory modeling for automatic recognition of speech," *Proceedings of the IEEE,* vol. 84, pp. 1199-1215, 1996.

[16] S. Deshpande, S. Chikkerur, and V. Govindaraju, "Accent classification in speech," in *Automatic Identification Advanced Technologies, 2005. Fourth IEEE Workshop on*, 2005, pp. 139-143.

[17] K. Liu Wai and P. Fung, "Fast accent identification and accented speech recognition," in *Acoustics, Speech, and Signal Processing, 1999. ICASSP '99. Proceedings., 1999 IEEE International Conference on*, 1999, pp. 221-224 vol.1.

[18] P. J. Ghesquiere and D. Van Compernolle, "Flemish accent identification based on formant and duration features," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, 2002, pp. I-749-I-752.

[19] M. A. Yusnita, M. P. Paulraj, S. Yaacob, S. A. Bakar, and A. Saidatul, "Malaysian English accents identification using LPC and formant analysis," in *2011 IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, 2011, pp. 472-476.

[20] H. Jue, L. Yi, T. F. Zheng, J. Olsen, and T. Jilei, "Multi-layered features with SVM for Chinese accent identification," in *Audio Language and Image Processing (ICALIP), 2010 International Conference on*, 2010, pp. 25-30.

[21] M. Do and M. Wagner, "Speaker recognition with small training requirements using a combination of VQ and DHMM," *Proc. of Speaker Recognition and Its Commercial and Forensic Applications,* pp. 169-172, 1998.

[22] M. Slaney, "Auditory Toolbox, Version 2, Technical Report No: 1998-010," Internal Research Corporation 1998.

[23] M. A. Yusnita, M. P. Paulraj, Y. Sazali, A. B. Shahriman, and Y. Rihana, "Analysis of Accent-sensitive Words in Mel-Frequency Cepstral Coefficients for Speaker Accent Identification of Malaysian English " in *4th International Conference on Noise, Vibration and Comfort (NVC2012)*, Kuala Lumpur, 2012, pp. 120-125.

[24] S. Furui, *Digital speech processing, synthesis, and recognition* vol. 7: CRC, 2001.

[25] M. A. Yusnita, M. P. Paulraj, S. Yaacob, A. B. Shahriman, and S. K. Nataraj, "Speaker Accent Recognition through Statistical Descriptors of Mel-bands Spectral Energy and Neural Network Model " in *The 3th IEEE Conference on Sustainable Utilization and Development in Engineering and Technology (IEEE STUDENT 2012)* Kuala Lumpur, 2012.

[26] L. W. Chew, K. P. Seng, L. M. Ang, V. Ramakonar, and A. Gnanasegaran, "Audio-emotion recognition system using parallel classifiers and audio feature analyzer," in *Proceedings - CIMSim 2011: 3rd International Conference on Computational Intelligence, Modelling and Simulation*, 2011, pp. 210-215.

[27] M. Rosell, "An introduction to front-end processing and acoustic features for automatic speech recognition," in *Lecture Notes of School of Computer Science and Communication, KTH, Sweden*, 2006.

[28] S. N. Sivanandam and M. P. Paulraj, *Introduction to Artificial Neural Networks* New Delhi: Vikas Publishing House PVT LTD, 2005.

[29] S. E. Fahlman, "An empirical study of learning speed in back-propagation networks," Carnegie Mellon University CMU-CS-88-162, 1 September 1988.

Yusnita Mohd Ali received her master degree in Electronics System Design Engineering from University Science of Malaysia in 2004. She is also a student member IEEE since 2009. She completed her Bachelor Degree in Electrical & Electronics Engineering from the same university in 1998. Currently, she is pursuing Ph.D. study in Universiti Malaysia Perlis and at the same time she is a lecturer in Universiti Teknologi MARA Malaysia. Her field of interest is speech and accent recognition, signal processing and artificial intelligence.

Paulraj Murugesa Pandiyan received BE in Electrical and Electronics Engineering from Madras University (1983), Master of Engineering in Computer Science and Engineering (1991) as well as Ph.D. in Computer Science from Bharathiyar University (2001), India. He is currently working as an Associate Professor in the school of Mechatronic Engineering, University Malaysia Perlis, Malaysia. His research interests include Principle, Analysis and Design of Intelligent Learning Algorithms, Brain Machine Interfacing, Dynamic Human Movement Analysis, Fuzzy Systems, and Acoustic Applications. He has co-authored a book on neural networks and 250 contributions in international journals and conference papers. He is a member of IEEE, member of the Institute of Engineers (India) and a life member in the System Society of India.

Sazali Yaacob received his Ph.D. Degree in Automatic Control and System Engineering from University of Sheffield, UK in 1995. He is working as a Professor and appointed as Head of Intelligence Signal Processing Group in Universiti Malaysia Perlis. He is interested in human behavior modeling, automatic control system and smart satellite system. He has published more than 180 international/national conference papers and more than 66 journal papers, 7 book chapters and 4 academic books.

Shahriman Abu Bakar received the B.E, M.E and PhD degrees in mechanical engineering from Mie University, Mie, Japan. After graduating he joined NEC Semiconductors (Malaysia) as Factory Automation Manager in Research and Development to improve equipments performance and efficiency. His research interest is in human motion analysis, ergonomics and system design. He has won 20 medals in various national and international research competitions and has more than 40 contributions in international journals and conference papers.

Nor Fadzilah Mokhtar received her master degree in Electronics System Design Engineering from University Science of Malaysia in 2004. She completed her Bachelor Degree in Electrical & Electronics Engineering from the same university in 1998. Currently, she is a lecturer in Universiti Teknologi MARA Malaysia. Her field of interest is in advance control system, artificial intelligence and RFID

.