International Journal of Automotive and Mechanical Engineering (IJAME)
ISSN: 2229-8649 (Print); ISSN: 2180-1606 (Online); Volume 7, pp. 1053-1073, January-June 2013

DOI: http://dx.doi.org/10.15282/ijame.7.2012.21.0086

ANALYSIS OF ACCENT-SENSITIVE WORDS IN MULTI-RESOLUTION MEL-FREQUENCY CEPSTRAL COEFFICIENTS FOR CLASSIFICATION OF ACCENTS IN MALAYSIAN ENGLISH

Yusnita M.A.[1,2,*], Paulraj M.P.[2], Sazali Yaacob[2], R. Yusuf[1] and Shahriman A.B.[2]

[1]Faculty of Electrical Engineering, Universiti Teknologi MARA,
13500 Permatang Pauh, Penang, Malaysia
[*]Email: yusnita082@ppinang.uitm.edu.my
Phone: +60135049526; Fax: +6043822776
[2]School of Mechatronic Engineering, Universiti Malaysia Perlis,
02600 Ulu Pauh, Perlis, Malaysia

ABSTRACT

This paper investigates the most accent-sensitive words for Malaysian English (MalE) speakers in multi-resolution 13 Mel-frequency cepstral coefficients. A text-independent accent system was implemented using different numbers of Mel-filters to determine the optimal settings for this database. Then, text-dependent accent systems were developed to rank the most accent-sensitive words for MalE speakers according to the classification rates. Prior work has also been conducted to test the significance of the wordlist for both gender and accent factors, and to investigate any interaction between these two factors. Experimental results show that male speakers have a higher intensity of accent effects compared with female speakers by 3.91% on text-independent and 3.47% on text-dependent tasks. Another finding has proven that by selecting appropriate words that carry severe accent markers could improve the task of speaker accent classification. An improvement of at most 8.45% and 8.91% was achieved on the male and female datasets, respectively, following vocabulary selection.

Keywords: Malaysian English; accent classification; mel-frequency cepstral coefficients; *K*-nearest neighbors; factorial design; analysis of variance.

INTRODUCTION

Nowadays, speech mining has become an interesting subject in human-machine interaction and communication systems. It carries abundant information that can be used to analyze human characteristics, such as gender, age, emotion, health state, and so forth. Accent, without exception, is one of the important traits in human biometrics, which is termed as voiceprint in speaker recognition systems. Accent is defined as a systematic variation in pronunciation patterns due to the ethno-linguistic and cultural background of a speaker. Malaysian English (MalE) is colored by different pronunciations because it is influenced by various ethnic groups (Nair Venugopal, 2000) within the country, which complicates the structure in comparison with native English pronunciations such as British English.

Through our recording observations, some voiced sounds like /z/ sometimes was sounded unvoiced or was substituted with the voiced /ʤ/ by the Malaysian Chinese such as in the word *zero* spelt as /zɪə.rəʊ/ using the International Phonetic Alphabet symbols. In the word *bottom* spelt as /ˈbɒt.ᵊm/, the Malays tend to unaspirate the voiceless alveolar plosive /t/ and substitute it with the glotal stop /ʔ/, whereas the

Chinese sound more aspirated. On the other hand, Indians substitute that with the voiceless retroflex plosive /ʈ/. In short, these people naturally sound more towards their own mother tongues rather than trying to articulate like a native. This unique attribute is what makes accent a potential means for identifying a speakers' characteristics, such as their ethnicity, and cultural and social-linguistic background. Above all, accent is the most influential factor next to gender that causes speaker/speech variability (Arslan and Hansen, 1996) and requires serious treatment in speech and speaker recognition systems. Prior works investigating language accent problems in English (Arslan and Hansen, 1996; Deshpande et al., 2005; Tanabian and Goubran, 2005; Pedersen and Diederich, 2007; Ullah and Karray, 2008), Persian (Rabiee and Setayeshi, 2010), Chinese (Huang et al., 2001; Too et al., 2001; Hou et al., 2010), and Flemish (Ghesquiere and Compernolle, 2002) have been reported. As English is recognized and practiced widely as the world's international language, investigations of accent problems in Malaysian's speech are is intriguing, but still in their infancy.

An engineering approach has been sought through various digital signal processing methods to extract the salient features of accent, such as Mel-frequency cepstral coefficients, abbreviated to MFCC (Arslan and Hansen, 1996; Humphries et al., 1996; Teixeira et al., 1996; Nguyen et al., 2010; Vergyri et al., 2010), linear prediction coefficients, abbreviated to LPC (Teixeira et al., 1996; Tanabian and Goubran, 2005), formant frequencies and trajectories (Liu and Fung, 1999; Deshpande et al., 2005; Tanabian and Goubran, 2005), and pitch contours (Hou et al., 2010; Vergyri et al., 2010). MFCC uses a perceptually motivated scale known as the Mel scale, which arises from the psychophysical study of the human auditory system. It is a scale used to measure subjective pitch for each tone as perceived by human ears. Meanwhile, methods to classify accent fall into two types: statistical approaches, such as the hidden Markov model (Arslan and Hansen, 1996; Humphries et al., 1996; Teixeira et al., 1996; Liu and Fung, 1999), support vector machine (Pedersen and Diederich, 2007), and *K*-nearest neighbor (KNN); and the other approach is a model evolved from human brain intelligence, namely artificial neural networks (Tanabian et al., 2005; Rabiee and Setayeshi, 2010).

The aims of this paper are threefold. Firstly, to split the spectrum of speech frames into several bands and to obtain Mel-frequency cepstral parameters as summations of all bands. Different Mel incremental steps are investigated to determine the best resolution in the Mel scale for extracting basic sounds of speech (phonemes). Past researchers have used arbitrary Mel-frequency resolutions, such as using 20 filters in the work by Do and Wagner (1998), and 40 filters in work by Slaney (1998). This paper attempts to experiment with multi-resolution Mel-frequencies to determine the optimum settings that match the characteristics of the MalE database. The second aim is to establish an appropriate wordlist selection for accent-sensitive words and to investigation the interaction between accent and gender factors using factorial design and ANOVA to test the hypotheses. Finally, to identify those words that carry better accent markers to be used in future accent classification tasks, which are not found in any past research for MalE to the best of the authors' knowledge.

METHODOLOGY

Speech Database

Several recording sessions were conducted to elicit speech from MalE speakers of three main ethnic group: Malays, Chinese, and Indians. The tasks consisted of three sections. Section A comprised 52 isolated words that were properly selected from currently popular accent databases, such as the CU-Accent Corpus (Arslan and Hansen, 1996; Hansen et. al., 2010) from the University of Texas in Dallas, the Speech Accent Archive (Weinberger, 2011) from George Mason University, and Speech Under Simulated and Actual Stress from Duke University. Section B comprised 17 sentences, formulated by our research group, which consisted of the aforementioned accent-sensitive wordlist in Section A. Finally, Section C utilized the Stella paragraph available at the Speech Accent Archive website. However, for the analysis purposes of this paper, only 18 isolated words selected perceptually from Section A were taken for this study. Table 1 presents the wordlist that was utilized. Table 2 describes the details of the portion of the MalE database (developed by Intelligent Signal Processing group at the University Malaysia Perlis) used in this work.

Table 1. Wordlist in Malaysian accent database.

No	Isolated Word (IW)	No	Isolated Word (IW)
1	Aluminum	10	Pleasure
2	Better	11	Station
3	Bottom	12	Stella
4	Boy	13	Student
5	Bringing	14	Target
6	Brother	15	Thirty
7	Communication	16	Time
8	Destination	17	Would
9	Girl	18	Zero

Table 2. Malaysian accent distribution of speakers and speech utterances.

Accent	Gender	No. of Speakers	No. of utterances (N)
Malay	Male	16	1440
	Female	22	1980
	Total	36	3420
Chinese	Male	19	1705
	Female	15	1350
	Total	34	3055
Indian	Male	13	1170
	Female	12	1080
	Total	25	2250
Total	Male	48	4315
	Female	49	4410
	Total	97	8725

Each word was repeated five times by each speaker to increase the number of samples per speaker, to increase precision, and to provide an estimate of the experimental error per word and per speaker. This collection of utterances amounted to 8725 speech samples recorded from 97 volunteers. The speakers originated from various regions of the country and as such, they were influenced by the regional accents. Subjects were postgraduate students of the Universiti Malaysia Perlis aged between 18 and 30 years old. The recording was carried out in a semi-anechoic acoustic chamber using a handheld condenser, supercardioid and unidirectional microphone using a laptop computer sound card and the MATLAB program. The recorded background noise level in the room was 22 dB. The sampling rate and bit resolution were set to 16 kHz and 16 bps for normal high quality, as used in automatic speech recognition applications.

Extraction of Mel-frequency Cepstral Coefficients

A block diagram showing the steps involved in extracting the MFCC is depicted in Figure 1. The working principle of the MFCC processor is based on a set of filter banks constructed from several bandpass filters, in the form of triangular-shaped window functions (Davis and Mermelstein, 1980). Filters are spaced uniformly on a perceptually motivated scale. The bandwidths are set so that 50% overlap with each other or a drop of half the power is laid on the middle point between the centers of adjacent filters. The center frequencies in Hz are mapped from the Mel scale; the known variation of the human ear's critical bandwidths with frequency. The mapping formulas are shown in Eq. (1) (Picone, 1993).

$$f_{mel} = 2595 \log_{10}\left(1 + \frac{f}{700}\right) \tag{1}$$

where f_{mel} and f are the Mel and linear frequencies, respectively.

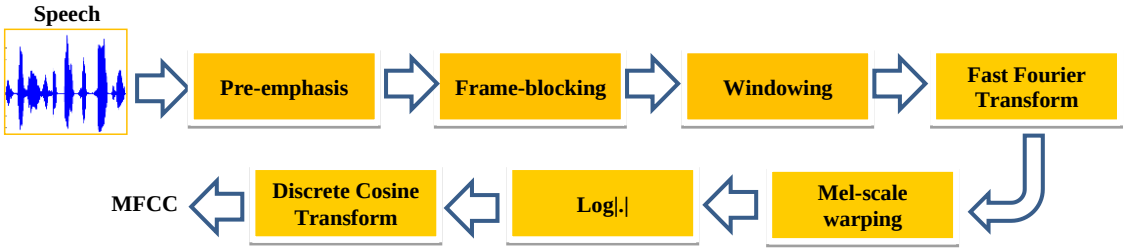


Figure 1. Block diagram of the MFCC feature extraction.

In this experiment, Mel-filters were designed based on the frequency range spanned by the filter banks, i.e., start and stop frequencies, FFT window length, the number of filters, N_{FB}, and the lower, center and upper frequencies. The latter three parameters were determined from the chosen sampling frequency and the number of filters. The increment step in the Mel scale can be calculated using Eq. (2), as in Plannerer (2004).

$$\Delta_{mel} = \frac{f_{mel(\max)}}{(N_{FB} + 1)} \tag{2}$$

where Δ_{mel}, $f_{mel(max)}$, and N_{FB} are the Mel step size, maximum Mel-frequency, and the number of Mel-filters in the filter bank, respectively.

The linearly spaced Mel-frequencies are converted to linear frequencies using Eq. (1), such that both scales are related almost linearly below 1 kHz, otherwise related logarithmically above 1 kHz. Figure 2 illustrates this mapping function. Figure 3 shows the resulting triangular filter banks residing on the linear frequency scale using an N_{FB} of 20. They are densely located for low frequencies, but sparsely located for higher frequencies. This infers that filtering using the Mel scale has emphasized the lower frequency components that are more important in speech analysis. The cepstral coefficients of the Mel-scale filter banks (Chew et al., 2011) can be computed, as by Eq. (3), by summing all the products of the fourier-transform-derived log-energy output of individual bandpass filters and discrete cosine transforms (DCT).

$$C_m = \sum_{k=1}^{N} E_k \cos\left[m(k-0.5)\pi / N\right] \tag{3}$$

where variables *C(.)* and *E(.)* represent the m^{th} cepstral coefficient (cepstrum) and the k^{th} log-energy, respectively. *N* is the number of filters in the filter banks and the number of the cepstrum is taken in the following order: i.e., *m=1, 2, 3, …,* M.

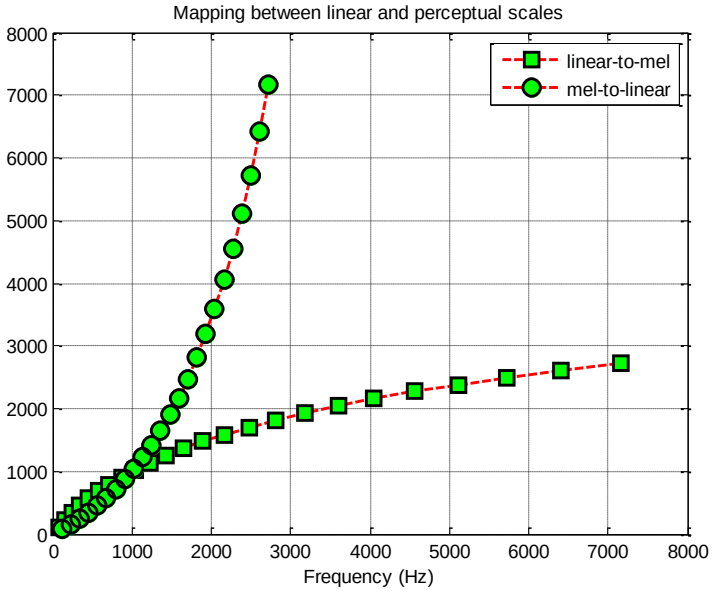


Figure 2. Mapping relationship of linear and Mel-frequency scales.

As the DCT performs the second frequency transform, the resulting new domain is a time-like domain called the frequency domain and the spectrum has become the cepstrum. The lower order cepstrum represents the slowly varying part of the spectrum, and spikes in the series correspond to the harmonic series of the vocal folds (Rosell, 2006). Normally, a few lower order coefficients are taken to represent the vocal tract shape, leaving out the pitch property of the speech signal.

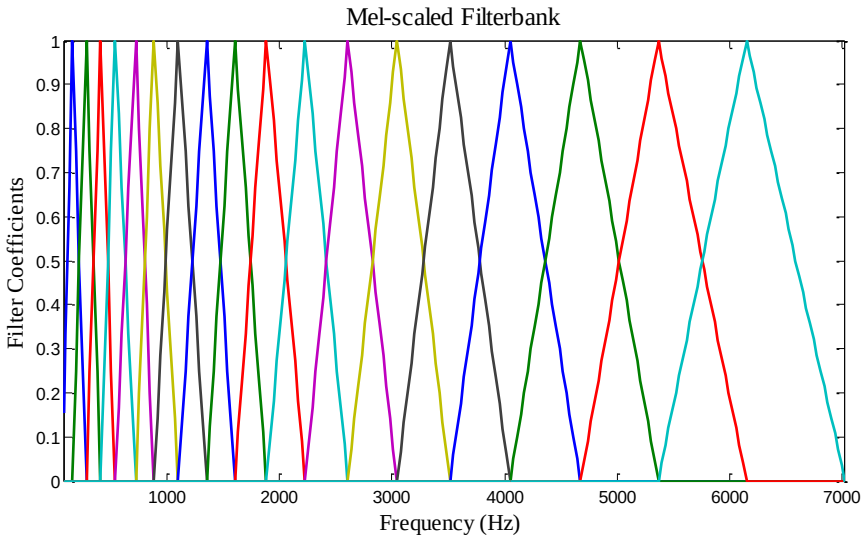


Figure 3. Mel filter banks basis functions using 20 Mel-filters in the filter bank.

Algorithm for MFCC Feature Extraction

Below is the formulated algorithm to explain the procedure of the MFCC extraction from the short-time windowed frames of the recorded acoustic signals.

Step 1: The sampled speech data are initially zero-adjusted to remove a DC bias during recording and to pre-emphasize using first-order FIR with a pre-emphasis parameter of 15/16 to compensate the attenuation in the spectral energy by approximately 6dB/octave.

Step 2: Frame-block the word samples into short time frames of 512 data points with 256 overlapping points and apply Hamming windows to the frames.

Step 3: Compute the spectrum from the pre-processed frames using FFT algorithm.

Step 4: In the frequency domain, sample each region of spectrum of interest by triangular-shaped windows, which are centered linearly in the Mel-scale, as in Eq. (1), using Mel-scale warped filters, as shown in Figure 3.

Step 5: Calculate the log of the spectral energy from the outputs of the Mel-filter banks resulted from *Step 4*.

Step 6: For each frame, compute the cepstral coefficients by applying DCT to all log-spectrum energies using Eq. (3). Take average values of the frames to form the feature vector.

Step 7: Repeat *Steps 1* to *6* for the other samples in the data collection and form a matrix of feature vectors (database) and assign class attributes to each pattern.

Factorial Design for Accent- and Gender-sensitive Words

This design of experiment (DOE) includes the simultaneous effects of multiple independent variables (factors) on a single dependent variable (response). Through this DOE, not only can the main effect of a single factor be studied, but also their interaction, if any. If significant interaction effects exist, the main effects have little practical meaning and therefore, they cannot be dropped as influential factors regardless

of their significance. It is required that at least each factor must have two levels and the treatments consist of the combinations of these levels. The number of runs required for each replicate is the product of the number of levels of all factors. In this paper, the effects of gender and accent on the extracted MFCC of acoustic signals (speech recordings) at different orders were investigated. The design of the study is specified in Table 3.

Figure 4 shows the arrangement of completely randomized design (CRD) for a 2-factor factorial design. Factor A (Gender) has $a = 2$ levels, factor B (Accent) has $b = 3$ levels, and $n = 5$ replicates. Each replication set contains all possible factor level combinations or treatments. The observations (values of the response variable) are defined as S_{ijm}^{k} where k is the MFCC-order of $k = 1, 2, 3, \ldots, 13$. The standard order will be randomized using CRD before taking any reading of the response variable. Figure 5 shows the sequence of standard order in each cell combination associated with S_{ijm}^{k}, as in Figure 4.

Table 3. The components of specification of the factorial design.

Component	Specification
Factor	Gender and accent
Type of factor	Both are classification factors
Treatment	Combinations of 2-level gender (male and female) and 3-level accent (Malay, Chinese, and Indian). Total number of combinations = 2 × 3 = 6 treatments in each replicate (base run)
Experimental unit	Speech (word utterance)
Replicate	5 speakers
Response variable	13-MFCC (each dependent variable of 1st order-MFCC, 2nd order-MFCC, 3rd order-MFCC, …, 13th order-MFCC)
Observation	Total run = N = 6 × 5 = 30 experiments to obtain the MFCC of each order

Factor A (Gender) \ Factor B (Accent)	1	2	3
1	S_{111}, S_{112}, S_{113}, S_{114}, S_{115}	S_{121}, S_{122}, S_{123}, S_{124}, S_{125}	S_{131}, S_{132}, S_{133}, S_{134}, S_{135}
2	S_{211}, S_{212}, S_{213}, S_{214}, S_{215}	S_{221}, S_{222}, S_{223}, S_{224}, S_{225}	S_{231}, S_{232}, S_{233}, S_{234}, S_{235}

Figure 4. Completely randomized design for 2-factor factorial design for any k^{th}-order of MFCC.

Factor A (Gender)	Factor B (Accent) 1	2	3
1	1, 7, 13, 19, 25	2, 8, 14, 20, 26	3, 9, 15, 21, 27
2	4, 10, 16, 22, 28	5, 11, 17, 23, 29	6, 12, 18, 24, 30

Figure 5. The sequence of original standard order of observations in completely randomized design treatment cells.

The hypotheses were made in testing the equality of means for different levels of factors on the MFCC scores using ANOVA, based on the following statistical (effects) model in Eq. (4). The hypotheses are given in Eq. (5).

$$S_{ijm} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijm} \begin{cases} i = 1, 2, \ \ldots, a \\ j = 1, 2, \ldots, b \\ m = 1, 2, \ldots, n \end{cases} \tag{4}$$

where *S(.)* is the statistical model for each MFCC-order, μ is an overall mean, τ_i is the effect of the i^{th} level of the row factor A, β_j is the effect of the j^{th} column of column factor B, $(\tau\beta)_{ij}$ is the interaction between τ_i and β_j, a random error term is defined as ε_{ijm}, and the subscript *m* indicates the replication index.

$$H_0: \ \tau_1 = \cdots = \tau_a = 0 \quad \text{v.s.} \quad H_1: \ \text{at least one } \tau_i \neq 0 \tag{5a}$$

$$H_0: \ \beta_1 = \cdots = \beta_b = 0 \quad \text{v.s.} \quad H_1: \ \text{at least one } \beta_j \neq 0 \tag{5b}$$

$$H_0: \ (\tau\beta)_{ij} = 0 \ \forall i, j \quad \text{v.s.} \quad H_1: \ \text{at least one } (\tau\beta)_{ij} \neq 0 \tag{5c}$$

_K_-nearest Neighbors Algorithm

The *K*-nearest neighbors (KNN) prediction of an unknown pattern, i.e., query instance, is based on a very simple majority vote of the categories or classes of the nearest neighbors in the training space. The underlying principle is based on minimum distances from the unlabeled sample to the training samples to determine the nearest *K*-neighbors. Euclidean distance is one of the popular methods used. This distance calculation from one sample or pattern in the testing dataset, which contains the unknown patterns, to one of the samples in the training dataset with known class labels, is expressed in Eq. (6).

$$d^2_{ij} = \sum_{m=1}^{M} \left[x_i(m) - x_j(m) \right]^2 \tag{6}$$

where $x_i(.)$ and $x_j(.)$ are exemplars of the training and the testing datasets in the m^{th} feature dimension, i.e., m=1, 2, …, M.

The next step is to locate the class number to the unlabeled pattern based on the majority vote (Tsang-Long et al., 2007) by simply summing up the class labels, assigned as $c(x_i)$where x_i is the class label of the selected $NK(x_j)$. The cardinality of $NK(x_j)$ is equal to K. Then, the subset of NN within the class set of $l \in \{1,2,...L\}$ is expressed mathematically as in Eq. (7).

$$N_K^l(x_j) = \{x_i \in NK(x_j) : c(x_i) = l\} \tag{7}$$

Thus, the classification result l^* using majority vote is expressed mathematically as in Eq. (8).

$$l^* = \arg_l^{\max} \left| N_K^l(x_j) \right| \tag{8}$$

The algorithm on how KNN works (Teknomo, 2011) is summarized as a flowchart in Figure 6. Normally, the *K*-parameter is determined by regression analysis. In general, it is chosen not to be a multiple integer of the number of classes, *L*.

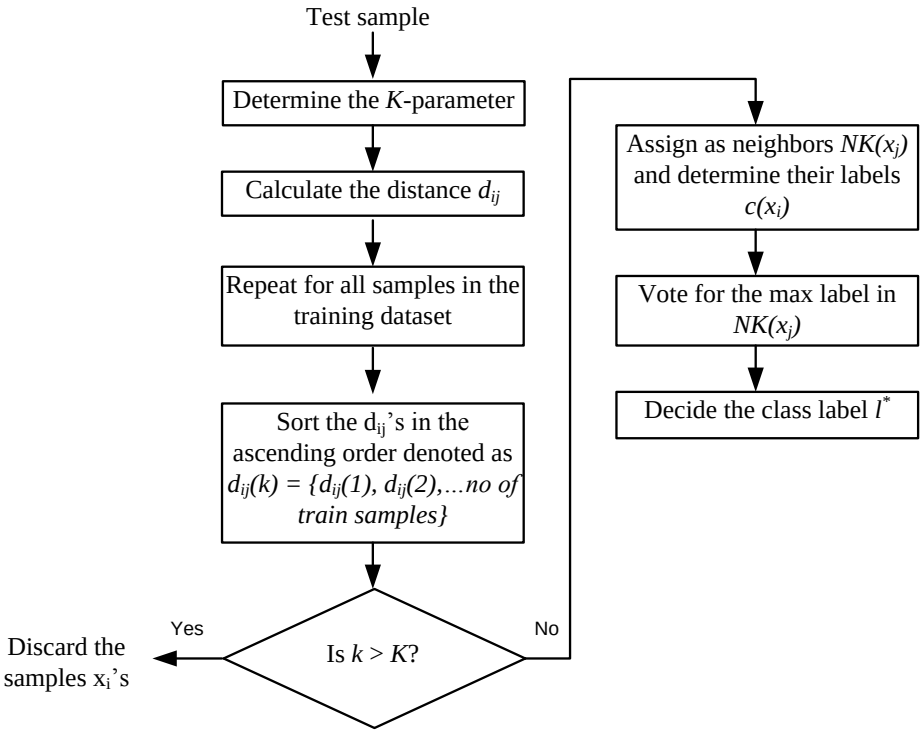


Figure 6. Flowchart of K-nearest neighbors algorithm.

RESULTS AND DISCUSSION

This section starts with the various designs of Mel-filter banks for the optimum setting of feature extraction. Next, prior works on establishing the appropriate wordlist selection for accent-sensitive words and investigations of the interaction between accent and gender factors, which are important for reducing the within group variability (inhomogeneous data), using factorial design and analysis of variance (ANOVA) to test the hypotheses are discussed. Rigorous analysis using several attempts of text-

independent and text-dependent MFCC-based accent classifier using the KNN model are reported here.

Mel-filter Banks Design

In order to investigate the effect of Mel-frequency scale resolution, the number of filters was varied from 40 to 10, in steps of 5, in order to obtain an optimum setting for this task. Table 4 tabulates the specifications of varying the number of filters N_{FB} in experimenting with different Mel resolutions. As there is little information below 150 Hz and above 6800 Hz for clean speech, for each design, some of the low- and high-ends of the filter banks were discarded in order to gain a twofold advantage, namely by getting rid of the 50-Hz hum from the AC power supply and by reducing the computational time.

Table 4. Various designs of Mel-filter banks.

No. of filters (N_{FB})	Max Mel freq (mel)	Mel resolution (mel)	Start freq. (Hz)	Stop freq. (Hz)	Omitted filters (N_{FB}^{th})
40	2840	69.3	141.7	6993.7	1, 2, 3, 39, & 40
35	2840	76.8	163.6	6863.4	1, 2, 3, 34, & 35
30	2840	91.6	193.3	7320.8	1, 2, 3, & 30
25	2840	109.2	149.7	7196.3	1, 2, & 25
20	2840	135.2	189.9	7016.2	1 & 20
15	2840	177.5	119.4	8000.0	1
10	2840	258.2	180.2	8000.0	1

Analysis of Factorial Design on Accent-sensitive Words

The analysis was partitioned into 18 different types of words uttered by 5 speakers (replicates), in order to identify words that are sensitive to accent and gender, and to establish the existence of interaction between these two factors. This prior work would establish a method to select the appropriate wordlist for accent classification and act as a basis to partition or not the overall experiments into different genders. Appendix A (A1–A4) shows the results of the ANOVA based on the statistical effects model and hypotheses given in Eqs. (4) and (5) for MFCC order k = 1 to 13. This number of coefficients has been used commonly in the past literature. For the purpose of factorial design, the number of filters N_{FB} was fixed at 30 as a control parameter. A summary of significant results of individual isolated words across each MFCC-order for accent factor is given in Table 5. The results of the interaction effect between gender and accent, as can be seen from Appendix A (A1–A4) are discussed. Out of the 18 words, only *Target* did not show any significant interaction effect at $p < 0.1$ for all MFCC orders. The others showed significant interaction effect at $p < 0.1$, for instance: *Destination*, *Girl*, *Pleasure*, *Time*, and *Zero* at k = 1, respectively; *Aluminum*, *Bottom*, *Boy*, *Brother*, *Communication*, *Stella*, and *Student* at k = 4, respectively; and *Better*, *Thirty*, *Would*, *Station*, and *Bringing* at k = 2, 3, 6, 9, and 10. These results suggest that

because the interaction effects for these wordlists were significant, the classification problem of accent should be partitioned into male and female.

Table 5 tabulates all the significant results for the main effects of accent factor at $p < 0.1$. The word *Better* has the greatest number of MFCC orders, which were found significant at $p < 0.1$. The lowest gains were recorded for the words *Brother* and *Destination*. On the other hand, across the wordlist, the 12th-order MFCC has shown the highest number of words that were significant, whereas the10th-order MFCC showed the least significant result, and the 8th-order MFCC has none that are significant across the wordlist. It was found that all words should be used in the next experiment, because the result has determined significance on the accent effect at the single order of MFCC.

Table 5. Summary of significant results of individual isolated words across MFCC-order for accent factor.

Isolated Word (IW)	Significance of mean difference (p < 0.1) of the accent factor for each MFCC order													
	1-MFCC	2-MFCC	3-MFCC	4-MFCC	5-MFCC	6-MFCC	7-MFCC	8-MFCC	9-MFCC	10-MFCC	11-MFCC	12-MFCC	13-MFCC	TOTAL
Aluminum			√	√	√							√		4
Better	√	√		√	√					√	√	√		7
Bottom		√	√		√								√	4
Boy		√											√	2
Bringing			√									√		2
Brother												√		1
Communication			√			√						√		3
Destination												√		1
Girl		√	√		√								√	4
Pleasure			√		√						√		√	4
Station						√						√		2
Stella						√	√		√					3
Student					√	√						√		3
Target	√		√	√					√			√		5
Thirty				√	√				√					3
Time			√	√			√							3
Would		√	√			√						√		4
Zero					√	√								2
TOTAL	2	5	9	5	8	6	2	0	3	1	2	10	4	

Figure 7 shows the factorial plots of the main effects and interaction effects of the word *Bottom* on the 5th-order MFCC and the word *Aluminum* on the 12th-order MFCC. It can be seen that both normality plots of residuals were normal in Figure 7(a) and (b); hence, the assumption of normality was not violated for ANOVA. For *Bottom* with the 5th-order MFCC, the interaction effect of gender and accent was significant, shown by the non-parallel plots in Figure 7(d) between the Chinese and Indian accents for different levels of gender. However, only the main effect of accent was significant as in Figure 7(c). For *Aluminum* with the 12th-order MFCC, the interaction effect between the factors was not significant, as is shown by Figure 7(f), because the difference in the response between the levels of accent was the same at all levels of gender. Here also, the main effect of gender was insignificant at p < 0.1, unlike the main effect of accent.

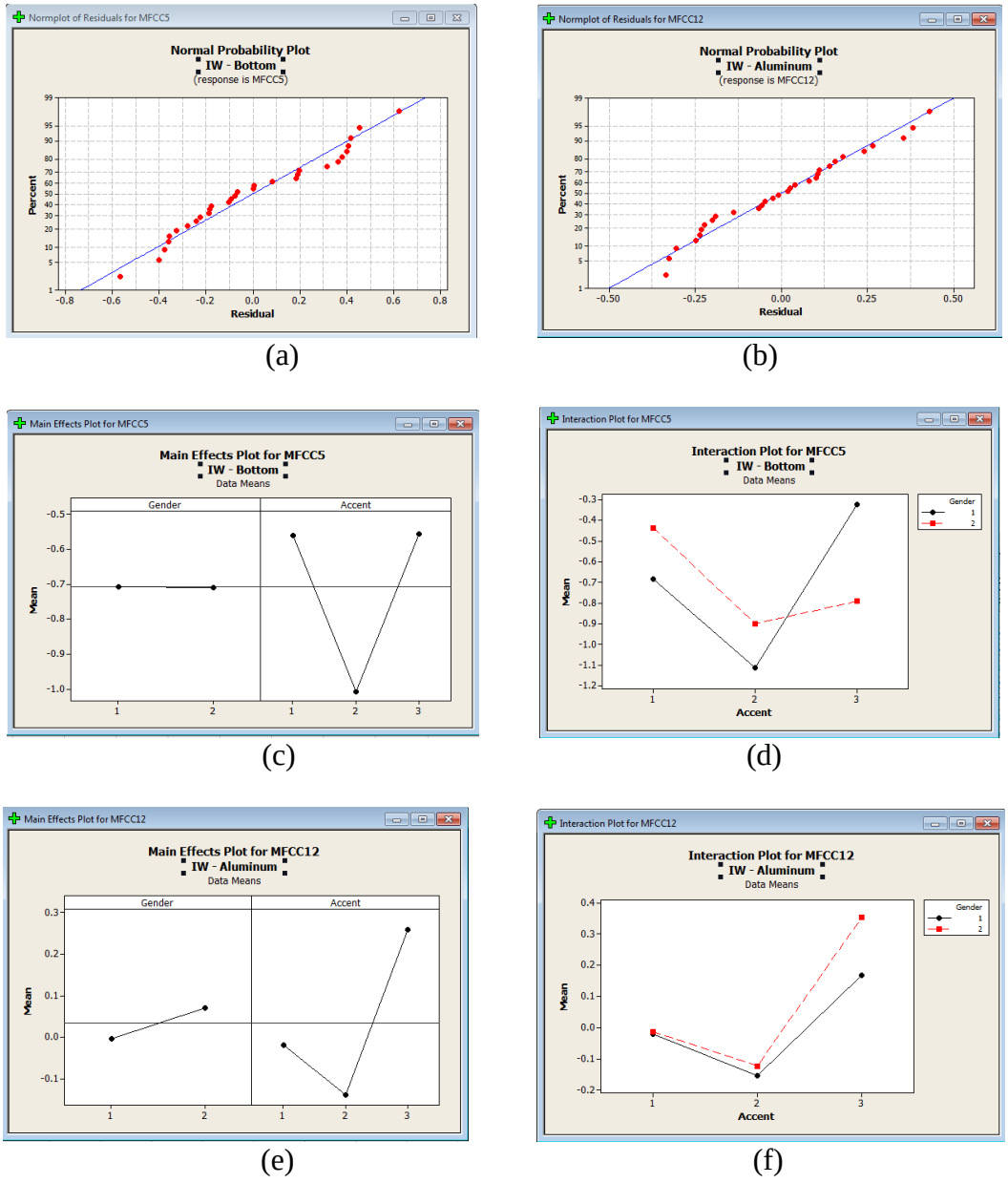


Figure 7. Normal plots of residual and factorial plots of main effects and interaction effects for the word *Aluminum* on 12th-order MFCC and the word *Bottom* on 5th-order MFCC.

Malaysian English Accent Classification

The KNN, as explained in the Methodology section, was used to model this system using varying numbers of *K*-parameter from 1 to 10. An independent test sample technique was implemented to evaluate the performance of this accent system using the parameterized 13-MFCC input features. The total database was reshuffled, randomized and partitioned into 70% as the training dataset and the remaining 30% of unseen data as the test dataset, separately for male and female datasets, owing to the existence of significant interaction between gender and accent, as found in the previous sub-section.

The performance was measured using classification rates (CRs) of correctly classified samples. All results reported in this paper were iterated and averaged over 10 trials.

Performance of Text-independent with Different Mel-filter Banks

For selecting optimal filters, 13-MFCC were derived from different number of Mel-filter banks, N_{FB} as in Table 4. The overall CRs resulting from using different N_{FB} are tabulated in Table 6. Only the results at K = 2 are recorded here. In general, both male and female datasets show an increase in CRs at higher numbers of N_{FB} with the highest CRs of the male and female speakers with an N_{FB} of 40 and 30 filters, respectively. It can be concluded that the choice of N_{FB} influenced the system performance. Using a smaller number of filters, such as $N_{FB} < 20$, would give too coarse resolution. For text-independent accent classification, the vocabulary set consisted of 18 words, as shown in Table 1. Next, the performance for the entire tests for different values of K-parameter, which were run over 10 trials each, is plotted in Figure 8(a) and (b) for the male and female speakers, respectively. Overall, the performance degraded at increasing K-value with K = 1 or 2 observed to be the best settings for this database. It seemed that much accent information has been lost when using the lower resolution of Mel-filter banks of N_{FB} = 10 and 15. The performance of text-independent accent classification for the male speakers was better than that of the female speakers by 3.91% at the highest recorded CRs of both.

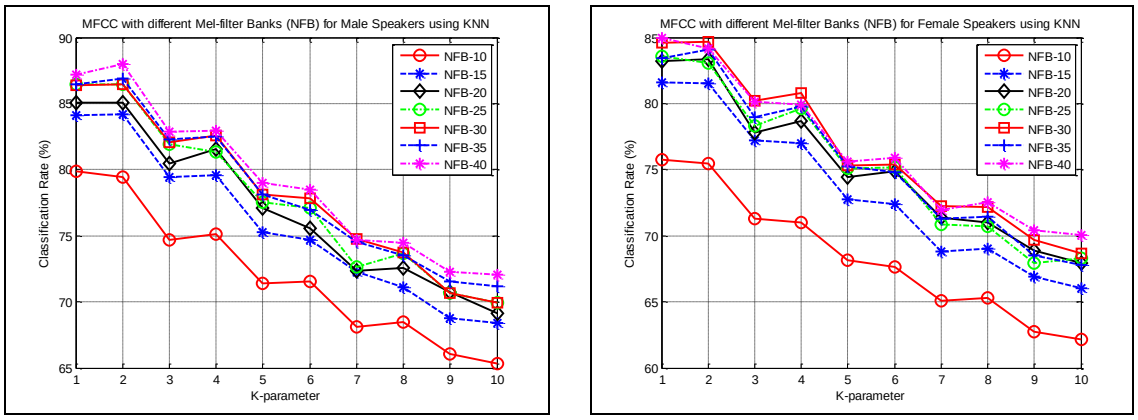
Table 6. Accuracy rates for text-independent accent classification across different filter numbers.

No. of filters (N_{FB})	Individual class classification rate (%)						Overall classification rate (%)	
	Malay		Chinese		Indian			
Gender	Male	Female	Male	Female	Male	Female	Male	Female
40	88.08	86.90	88.69	83.78	86.78	79.51	**87.97**	84.13
35	87.55	87.35	87.36	84.05	85.38	78.02	86.89	84.05
30	87.48	88.57	86.59	83.48	85.07	78.98	86.48	**84.66**
25	86.46	86.57	87.67	81.83	84.76	78.15	86.48	83.05
20	85.44	87.30	85.87	82.49	83.42	77.25	85.06	83.37
15	84.14	85.34	85.46	79.70	82.34	76.76	84.17	81.51
10	80.23	80.02	80.61	74.67	76.70	68.18	79.42	75.48

Performance of Text-dependent with Fixed Mel-filter Banks

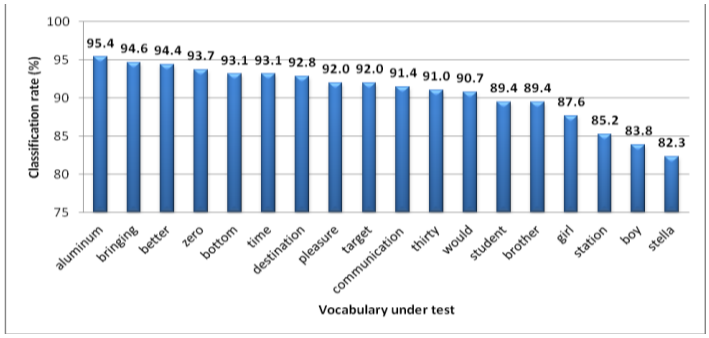
For testing the intensity of accent-sensitive words, N_{FB} was fixed at 40 and 30 filters for the male and female datasets, respectively, as obtained from the previous experiments. For text-dependent accent classification, the vocabulary set consisted of single word, speaker independent, trained and tested separately using KNN by varying K = 1 to 10. The best results were taken for plotting, which occurred only at K = 1 or 2. Figure 9(a) and (b) shows the bar chart rankings of the highest to lowest accent-sensitive words. It was found that different genders experienced different intensity of accent detection and different ranking of most accent-sensitive words. For example, the first five most accent-sensitive words for the male speakers were *aluminum*, *bringing*, *better*, *zero*, and *bottom*; whereas for the female speakers they were *target*, *destination*, *would*, *bottom*, and *girl*. Figure 9(a) shows a larger difference in the CRs of 13.1% between the two

extreme values, whereas Figure 9(b) shows a smaller difference in the CRs of 8.5% between the two extreme values. Most of the words in the middle ranking shared similar results for the female dataset. However, the male dataset exhibited noticeable differences. On the highest word rate, the male speakers outperformed the female speakers by the CR of 3.47%.

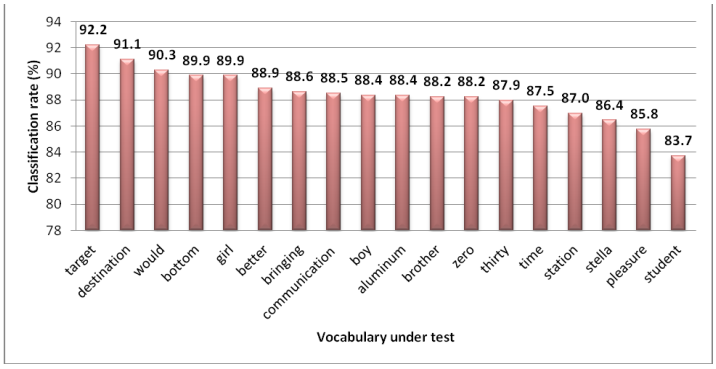


(a) (b)

Figure 8. Plots of the performance of accent classifiers using different Mel-filter banks (N_{FB}) in MFCC features for: (a) male speakers, and (b) female speakers across different K-parameters.



(a)



(b)

Figure 9. Vocabulary ranking in terms of accent accuracy rates for: (a) male speakers, and (b) female speakers using the KNN classifier and independent test samples with a partition of 70% training and 30% test.

CONCLUSION

This paper has presented a study of accent-sensitive words in the MFCC-based feature vector space and Mel-frequency resolution analysis using different numbers of filters. Prior work to establish the appropriate wordlist selection for accent-sensitive words and the investigation of the interaction between accent and gender factors was conducted using factorial design and ANOVA to test the hypotheses. Both text-independent and text-dependent accent classification were implemented using the KNN classification algorithm and the performances of the system were evaluated using the independent test samples technique. The best overall accuracy rates of 87.97% and 84.66% were obtained using 40 and 30 filters, respectively, for the male and female datasets for text-independent accent system, which included all the words that were determined significant to accent from the earlier ANOVA tests. Following this, text-dependent systems on individual isolated words were conducted to rank accent-sensitive words according to gender. It was found that the male speakers demonstrated higher intensity of accent effects compared with the female speakers, by 3.91% on text-independent tasks and by 3.47% on text-dependent tasks on the selected best results. From the experiments conducted, it was proven that selecting appropriate words that carry severe accent markers works satisfactorily in the task of speaker accent classification. The improvement was made by at most 8.45% and 8.91%, respectively for the male and female datasets, following vocabulary selection.

ACKNOWLEDGMENTS

The authors would like to acknowledge the encouragement given by the Vice Chancellor of the University Malaysia Perlis (UniMAP), Brig. Jeneral Dato' Prof. Dr. Kamaruddin Hussin, and the financial assistance under fundamental research grant scheme numbered 9003- 00278 and also the sponsorship of the PhD candidature from the Ministry of Higher Education, Malaysia.

REFERENCES

Arslan, L.M. and Hansen, J.H.L. 1996. Language accent classification in American English. Speech Communication, 18(4): 353-367.

Chew, L.W., Seng, K.P., Ang, L.M., Ramakonar, V. and Gnanasegaran, A. 2011. Audio-emotion recognition system using parallel classifiers and audio feature analyzer. Proceedings of the 3rd International Conference on Computational Intelligence, Modelling and Simulation (CIMSim 2011), pp. 210-215.

Davis, S. and Mermelstein, P. 1980. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Transactions on Acoustics, Speech and Signal Processing, 28(4): 357-366.

Deshpande, S., Chikkerur, S. and Govindaraju, V. 2005. Accent classification in speech. Proceedings of the Fourth IEEE Workshop on the Automatic Identification Advanced Technologies, pp. 139-143.

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

Ghesquiere, P.J. and Compernolle, D.V. 2002. Flemish accent identification based on formant and duration features. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. I-749-I-752.

Hansen, J.H.L., Gray, S.S. and Kim, W. 2010. Automatic voice onset time detection for unvoiced stops (/p/,/t/,/k/) with application to accent classification. Speech Communication, 52(10): 777-789.

Hou, J., Liu, Y., Zheng, T.F., Olsen, J. and Tian, J. 2010. Multi-layered features with SVM for Chinese accent identification. Proceedings of International Conference on Audio Language and Image Processing (ICALIP), pp. 25-30.

Huang, C., Chen, T., Li, S., Chang, E. and Zhou, J. 2001. Analysis of speaker variability. Proceedings of Eurospeech 2001, pp. 1377–1380.

Humphries, J.J., Woodland, P.C. and Pearce, D. 1996. Using accent-specific pronunciation modelling for robust speech recognition. Proceedings of the 4th International Conference on Spoken Language (ICSLP 1996), pp. 2324-2327.

Liu, W.K. and Fung, P. 1999. Fast accent identification and accented speech recognition. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 1999), pp. 221-224.

Nair Venugopal, S. (2000). English, identity and the Malaysian workplace. World Englishes, 19: 205-213.

Nguyen, P., Tran, D., Huang, X. and Sharma, D. 2010. Australian accent-based speaker classification. Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining (WKDD 2010), pp. 416-419.

Pedersen, C. and Diederich, J. 2007. Accent Classification Using Support Vector Machines. Proceedings of the 6th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2007), pp. 444-449.

Picone, J. W. 1993. Signal modeling techniques in speech recognition. Proceedings of the IEEE, 81(9): 1215-1247.

Plannerer, B. 2004. Speech Recognition Workshop Exercises Using MATLAB. http://www.speech-recognition.de/matlab-examples.html. (Accessed on 15 January 2012).

Rabiee, A. and Setayeshi, S. 2010. Persian accents identification using an adaptive neural network. Proceedings of the 2nd International Workshop on Education Technology and Computer Science (ETCS 2010), pp. 7-10.

Rosell, M. 2006. An introduction to front-end processing and acoustic features for automatic speech recognition. http://www.nada.kth.se/~rosell/courses/ rosell_acoustic_features.pdf. (Accessed on 23 December 2012).

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

Tanabian, M.M. and Goubran, R.A. 2005. Speech accent identification with vocal tract variation trajectory tracking using neural networks. Proceedings of the 2005 IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety (CIHSPS 2005), pp. 117-121.

Tanabian, M.M., Tierney, P. and Azami, B.Z. 2005. Automatic speaker recognition with formant trajectory tracking using CART and neural networks. Proceedings of the Canadian Conference on Electrical and Computer Engineering, pp. 1225-1228.

Teixeira, C., Trancoso, I. and Serralheiro, A. 1996. Accent identification. Proceedings of the 4th International Conference on Spoken Language (ICSLP 1996), pp. 1784-1787.

Teknomo, K. 2011. K-Nearest Neighbors Tutorial. http:\\people.revoledu.com\kardi\ tutorial\KNN\. (Accessed on 12 Mac 2012).

Too, C., Chao, H., Chang, E. and Jingehan, W. 2001. Automatic accent identification using Gaussian mixture models. Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU 2001), pp. 343-346.

Tsang-Long, P., Wen-Yuan, L. and Yu-Te, C. 2007. Audio-Visual Speech Recognition with Weighted KNN-based Classification in Mandarin Database. Proceedings of the 3rd International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP 2007), pp. 39-42.

Ullah, S. and Karray, F. 2008. An evolutionary approach for accent classification in IVR systems. Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC 2008), pp. 418-423.

Vergyri, D., Lamel, L. and Gauvain, J. L. 2010. Automatic Speech Recognition of Multiple Accented English Data. Proceedings of the INTERSPEECH 2010, pp. 1652-1655.

Weinberger, S.H. 2011. The Speech Accent Archive. http://accent.gmu.edu/. (Accessed on 10 January 2012).

APPENDIX

A1. Results of ANOVA on different orders of MFCC scores (1st- to 4th-order)

Isolated Word (IW)	p-value for the 1st order MFCC					p-value for the 2nd order MFCC				
	Gen	Acc	Gen∗ Acc	R-Sq (%)	R-Sq Adjusted (%)	Gen	Acc	Gen∗ Acc	R-Sq (%)	R-Sq Adjusted (%)
Aluminum	0.804	0.215	0.557	15.91	0.00	0.035♣	0.189	0.850	27.10	11.91
Better	0.440	0.010♣	0.634	34.56	20.93	0.984	0.173	0.043♣	31.42	17.13
Bottom	0.287	0.914	0.611	9.00	0.00	0.416	0.023♣	0.428	32.14	18.00
Boy	0.698	0.862	0.115	17.78	0.660	0.574	0.039♣	0.407	28.60	13.73
Bringing	0.109♣	0.635	0.176	23.63	7.72	0.097♣	0.178	0.626	24.20	8.41
Brother	0.001♣	0.549	0.178	44.25	32.63	0.001♣	0.719	0.923	39.52	26.92
Communication	0.590	0.621	0.294	13.82	0.00	0.104♣	0.717	0.634	15.65	0.00
Destination	0.922	0.588	0.035♣	26.84	11.60	0.378	0.114	0.261	25.95	10.52
Girl	0.199	0.299	0.074♣	29.62	14.95	0.959	0.017♣	0.290	33.79	20.00
Pleasure	0.474	0.913	0.088♣	20.24	3.62	0.008♣	0.706	0.783	28.80	13.96
Station	0.433	0.946	0.161	16.35	0.00	0.956	0.218	0.116	24.95	9.32
Stella	0.022♣	0.716	0.238	28.84	14.01	0.089♣	0.436	0.602	19.70	2.97
Student	0.388	0.215	0.755	16.15	0.00	0.628	0.202	0.775	14.83	0.00
Target	0.910	0.081♣	0.922	19.36	2.56	0.421	0.327	0.243	20.03	3.36
Thirty	0.272	0.903	0.546	10.15	0.00	0.742	0.210	0.956	12.84	0.00
Time	0.806	0.798	0.0001♣	48.18	37.38	0.190	0.174	0.377	24.08	8.27
Would	0.238	0.195	0.747	18.79	1.88	0.432	0.096♣	0.656	21.75	5.45
Zero	0.466	0.516	0.059♣	25.64	10.15	0.590	0.149	0.289	22.67	6.56
Isolated Word (IW)	p-value for the 3rd order MFCC					p-value for the 4th order MFCC				
	Gen	Acc	Gen∗ Acc	R-Sq (%)	R-Sq Adjusted (%)	Gen	Acc	Gen∗ Acc	R-Sq (%)	R-Sq Adjusted (%)
Aluminum	0.065♣	0.034♣	0.235	37.91	24.98	0.097♣	0.044♣	0.015♣	45.54	34.20
Better	0.868	0.006♣	0.934	35.04	21.51	0.034♣	0.124	0.001♣	53.13	43.36
Bottom	0.486	0.031♣	0.836	27.03	11.83	0.0001♣	0.122	0.003♣	63.23	55.57
Boy	0.724	0.249	0.928	11.85	0.00	0.0001♣	0.129	0.029♣	61.55	53.54
Bringing	0.012♣	0.075♣	0.731	36.54	23.32	0.427	0.428	0.403	15.19	0.00
Brother	0.274	0.268	0.982	14.51	0.00	0.007♣	0.148	0.005♣	52.33	42.40
Communication	0.029♣	0.044♣	0.224	39.60	27.02	0.008♣	0.210	0.068♣	42.39	30.39
Destination	0.004♣	0.250	0.030♣	46.91	35.85	0.358	0.220	0.087♣	28.38	13.46
Girl	0.366	0.050♣	0.543	27.02	11.82	0.002♣	0.221	0.467	41.30	29.07
Pleasure	0.074♣	0.003♣	0.919	43.91	32.23	0.126	0.346	0.458	20.90	4.42
Station	0.007♣	0.908	0.137	35.68	22.28	0.964	0.444	0.664	9.49	0.00
Stella	0.241	0.301	0.857	15.14	0.00	0.101♣	0.362	0.011♣	39.86	27.36
Student	0.037♣	0.134	0.483	30.89	16.49	0.371	0.717	0.009♣	35.13	21.61
Target	0.939	0.034♣	0.781	25.70	10.22	0.348	0.085♣	0.212	28.82	13.99
Thirty	0.860	0.638	0.051♣	24.26	8.48	0.446	0.043♣	0.009♣	44.67	33.08
Time	0.783	0.045♣	0.545	25.97	10.55	0.722	0.025♣	0.083♣	37.25	24.18
Would	0.323	0.061♣	0.632	25.60	10.10	0.180	0.703	0.581	13.45	0.00
Zero	0.225	0.216	0.968	16.91	0.00	0.002♣	0.104	0.052♣	49.09	38.48

A2. Results of ANOVA on different orders of MFCC scores (5th- to 8th-order)

Isolated Word (IW)	p-value for the 5th order MFCC					p-value for the 6th order MFCC				
	Gen	Acc	Gen∗ Acc	R-Sq (%)	R-Sq Adjusted (%)	Gen	Acc	Gen∗ Acc	R-Sq (%)	R-Sq Adjusted (%)
Aluminum	0.076♣	0.006♣	0.045♣	49.49	38.96	0.508	0.941	0.561	6.83	0.00
Better	0.101♣	0.001♣	0.040♣	53.90	44.29	0.048♣	0.026♣	0.710	36.18	22.88
Bottom	0.984	0.010♣	0.049♣	42.88	30.98	0.114	0.247	0.288	25.64	10.15
Boy	0.009♣	0.766	0.254	32.37	18.28	0.011♣	0.260	0.096♣	39.57	26.98
Bringing	0.0001♣	0.201	0.755	58.85	50.28	0.001♣	0.464	0.272	43.28	31.46
Brother	0.704	0.121	0.236	24.64	8.94	0.130	0.191	0.863	20.81	4.31
Communication	0.0001♣	0.265	0.540	54.48	45.00	0.093♣	0.046♣	0.485	32.54	18.49
Destination	0.0001♣	0.171	0.278	65.99	58.51	0.642	0.564	0.636	8.80	0.00
Girl	0.001♣	0.066♣	0.987	44.79	33.29	0.974	0.861	0.903	2.07	0.00
Pleasure	0.0001♣	0.012♣	0.036♣	60.16	51.86	0.150	0.474	0.481	17.98	0.89
Station	0.0001♣	0.115	0.670	59.04	50.51	0.186	0.021♣	0.565	33.54	19.70
Stella	0.836	0.162	0.943	14.58	0.00	0.612	0.055♣	0.598	24.69	9.00
Student	0.034♣	0.012♣	0.055♣	48.14	37.34	0.652	0.103♣	0.932	18.27	1.24
Target	0.008♣	0.760	0.197	33.94	20.17	0.115	0.154	0.887	22.49	6.34
Thirty	0.0001♣	0.035♣	0.290	57.31	48.42	0.581	0.148	0.160	25.96	10.53
Time	0.056♣	0.682	0.082♣	30.22	15.69	0.004♣	0.596	0.453	34.77	21.18
Would	0.003♣	0.534	0.859	33.65	19.83	0.001♣	0.010♣	0.103♣	55.20	45.86
Zero	0.0001♣	0.004♣	0.603	56.58	47.53	0.075♣	0.041♣	0.141	38.48	25.67
Isolated Word (IW)	p-value for the 7th order MFCC					p-value for the 8th order MFCC				
	Gen	Acc	Gen∗ Acc	R-Sq (%)	R-Sq Adjusted (%)	Gen	Acc	Gen∗ Acc	R-Sq (%)	R-Sq Adjusted (%)
Aluminum	0.207	0.706	0.484	13.92	0.00	0.011♣	0.862	0.234	31.38	17.08
Better	0.261	0.766	0.777	9.00	0.00	0.0001♣	0.173	0.725	65.00	57.71
Bottom	0.126	0.755	0.392	17.33	0.11	0.177	0.145	0.883	20.98	4.52
Boy	0.074♣	0.216	0.688	23.83	7.96	0.001♣	0.293	0.864	44.00	32.33
Bringing	0.480	0.372	0.267	18.29	1.26	0.905	0.821	0.840	3.08	0.00
Brother	0.013♣	0.319	0.783	29.42	14.72	0.0001♣	0.800	0.064♣	70.34	64.16
Communication	0.953	0.201	0.686	14.89	0.00	0.005♣	0.798	0.608	31.36	17.06
Destination	0.153	0.448	0.309	20.82	4.32	0.0001♣	0.575	0.364	60.78	52.61
Girl	0.003♣	0.292	0.061♣	45.21	33.80	0.003♣	0.401	0.160	41.15	28.88
Pleasure	0.136	0.938	0.663	12.24	0.00	0.0001♣	0.574	0.063♣	70.24	64.04
Station	0.850	0.724	0.471	8.55	0.00	0.0001♣	0.578	0.773	60.67	52.48
Stella	0.665	0.060♣	0.529	24.59	8.88	0.0001♣	0.905	0.602	45.47	34.11
Student	0.008♣	0.740	0.549	29.91	15.31	0.0001♣	0.315	0.430	50.06	39.65
Target	0.170	0.702	0.488	14.90	0.00	0.0001♣	0.811	0.143	52.85	43.02
Thirty	0.076♣	0.740	0.235	22.90	6.84	0.003♣	0.208	0.257	41.54	29.36
Time	0.122	0.026♣	0.928	31.99	17.82	0.001♣	0.417	0.383	44.00	32.33
Would	0.0001♣	0.582	0.882	48.85	38.19	0.175	0.586	0.953	11.58	0.00
Zero	0.071♣	0.128	0.572	27.68	12.61	0.061♣	0.396	0.831	20.47	3.90

A3. Results of ANOVA on different orders of MFCC scores (9th- to 12th-order)

Isolated Word (IW)	p-value for the 9th order MFCC					p-value for the 10th order MFCC				
	Gen	Acc	Gen∗ Acc	R-Sq (%)	R-Sq Adjusted (%)	Gen	Acc	Gen∗ Acc	R-Sq (%)	R-Sq Adjusted (%)
Aluminum	0.005♣	0.593	0.784	32.07	17.92	0.0001♣	0.947	0.151	55.60	46.35
Better	0.0001♣	0.357	0.436	50.78	40.52	0.0001♣	0.081♣	0.314	70.61	64.48
Bottom	0.921	0.893	0.561	5.59	0.00	0.0001♣	0.547	0.955	53.04	43.26
Boy	0.344	0.686	0.539	11.00	0.00	0.0001♣	0.294	0.956	46.27	35.08
Bringing	0.001♣	0.435	0.814	38.78	26.02	0.0001♣	0.208	0.049♣	63.81	56.28
Brother	0.0001♣	0.404	0.484	54.40	44.90	0.0001♣	0.345	0.225	59.31	50.83
Communication	0.0001♣	0.931	0.912	57.98	49.22	0.0001♣	0.943	0.597	43.27	31.45
Destination	0.016♣	0.545	0.089♣	35.66	22.25	0.028♣	0.170	0.040♣	40.99	28.70
Girl	0.350	0.468	0.612	12.66	0.00	0.0001♣	0.177	0.178	62.68	54.90
Pleasure	0.269	0.365	0.108♣	25.64	10.15	0.0001♣	0.266	0.496	60.88	52.73
Station	0.180	0.461	0.076♣	27.82	12.79	0.543	0.208	0.625	16.36	0.00
Stella	0.0001♣	0.022♣	0.176	57.10	48.17	0.005♣	0.777	0.667	30.72	16.29
Student	0.077♣	0.977	0.535	16.51	0.00	0.005♣	0.230	0.017♣	48.49	37.76
Target	0.602	0.053♣	0.207	30.06	15.49	0.0001♣	0.377	0.337	59.96	51.62
Thirty	0.010♣	0.098♣	0.170	41.09	28.81	0.0001♣	0.407	0.262	50.51	40.19
Time	0.582	0.134	0.041♣	33.32	19.43	0.040♣	0.940	0.670	19.04	2.17
Would	0.025♣	0.757	0.580	23.45	7.50	0.120	0.187	0.433	24.85	9.20
Zero	0.233	0.578	0.369	16.38	0.00	0.004♣	0.768	0.209	37.35	24.29
Isolated Word (IW)	p-value for the 11th order MFCC					p-value for the 12th order MFCC				
	Gen	Acc	Gen∗ Acc	R-Sq (%)	R-Sq Adjusted (%)	Gen	Acc	Gen∗ Acc	R-Sq (%)	R-Sq Adjusted (%)
Aluminum	0.231	0.374	0.401	18.54	1.57	0.395	0.003♣	0.66	40.86	28.54
Better	0.495	0.027♣	0.123	36.00	22.67	0.220	0.015♣	0.605	34.46	20.80
Bottom	0.861	0.777	0.700	5.01	0.00	0.684	0.138	0.735	17.52	0.34
Boy	0.0001♣	0.877	0.198	46.13	34.90	0.625	0.548	0.882	6.73	0.00
Bringing	0.020♣	0.398	0.357	29.90	15.29	0.498	0.118	0.719	19.54	2.78
Brother	0.006♣	0.558	0.829	31.05	16.69	0.103♣	0.006♣	0.403	42.11	30.05
Communication	0.001♣	0.899	0.118	42.74	30.81	0.009♣	0.107♣	0.882	35.69	22.29
Destination	0.061♣	0.484	0.165	27.79	12.75	0.004♣	0.004♣	0.617	51.40	41.28
Girl	0.493	0.660	0.164	17.90	0.79	0.619	0.455	0.981	7.40	0.00
Pleasure	0.898	0.059♣	0.228	28.50	13.60	0.005♣	0.422	0.992	32.26	18.15
Station	0.665	0.804	0.079♣	20.73	4.21	0.004♣	0.011♣	0.618	48.22	37.43
Stella	0.003♣	0.338	0.779	36.25	22.97	0.131	0.438	0.983	14.84	0.00
Student	0.225	0.154	0.014♣	39.69	27.13	0.275	0.058♣	0.842	25.00	9.38
Target	0.583	0.930	0.723	4.43	0.00	0.542	0.016♣	0.547	32.50	18.43
Thirty	0.255	0.528	0.678	12.61	0.00	0.667	0.419	0.854	8.80	0.00
Time	0.362	0.953	0.866	4.96	0.00	0.165	0.122	0.444	29.76	10.29
Would	0.001♣	0.493	0.313	44.59	33.05	0.983	0.070♣	0.139	29.91	15.31
Zero	0.0001♣	0.274	0.501	61.82	53.86	0.179	0.201	0.275	25.20	9.61

A4. Results of ANOVA on different orders of MFCC scores (13th-order)

Isolated Word (IW)	p-value for the 13th order MFCC				
	Gen	Acc	Gen∗ Acc	R-Sq (%)	R-Sq Adjusted (%)
Aluminum	0.0001♣	0.657	0.040♣	62.75	54.98
Better	0.0001♣	0.271	0.365	59.69	51.30
Bottom	0.0001♣	0.041♣	0.237	67.45	60.67
Boy	0.0001♣	0.028♣	0.514	54.00	44.42
Bringing	0.0001♣	0.573	0.234	73.40	67.86
Brother	0.0001♣	0.247	0.056♣	80.16	76.03
Communication	0.0001♣	0.695	0.359	72.89	67.24
Destination	0.0001♣	0.157	0.821	65.49	58.30
Girl	0.0001♣	0.032♣	0.0001♣	84.41	81.16
Pleasure	0.0001♣	0.087♣	0.208	80.40	76.31
Station	0.001♣	0.773	0.739	37.47	24.44
Stella	0.0001♣	0.215	0.451	72.57	66.85
Student	0.0001♣	0.783	0.221	76.85	72.02
Target	0.0001♣	0.659	0.311	78.32	73.81
Thirty	0.0001♣	0.706	0.227	49.61	39.11
Time	0.001♣	0.294	0.676	43.85	32.15
Would	0.0001♣	0.503	0.021♣	77.72	73.08
Zero	0.0001♣	0.524	0.017♣	79.22	74.89

♣The mean difference is significant at the 0.10 level ($p < 0.1$)
∗Interaction