IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 15 April 2023, accepted 7 July 2023, date of publication 28 July 2023, date of current version 7 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3299814

RESEARCH ARTICLE

The Study of Malay's Prosodic Features Impact on Classical Arabic Accents Recognition

NOOR JAMALIAH IBRAHIM[1,4], MOHD YAMANI IDNA IDRIS[1,4], (Member, IEEE), M. Y. ZULKIFLI MOHD YUSOFF[2,4], ROZIANA RAMLI[1], AND RAJA JAMILAH RAJA YUSOF[3,4], (Senior Member, IEEE)

[1]Department of Computer System and Technology, Faculty of Computer Science and Information Technology, Universiti Malaya, Kuala Lumpur 50603, Malaysia
[2]Department of Al-Quran and Al-Hadith, Academy of Islamic Studies, Universiti Malaya, Kuala Lumpur 50603, Malaysia
[3]Department of Software Engineering, Faculty of Computer Science and Information Technology, Universiti Malaya, Kuala Lumpur 50603, Malaysia
[4]Centre of Quranic Research (CQR), Universiti Malaya, Kuala Lumpur 59990, Malaysia

Corresponding authors: Noor Jamaliah Ibrahim (noor3184.um@gmail.com) and Mohd Yamani Idna Idris (yamani@um.edu.my)

ABSTRACT Modeling individual's variation in speech pattern can be challenging in Automatic Speech Recognition (ASR). In Classical Arabic (CA) language, 20 Quranic accents are permitted for Quranic recitation. An ASR system for CA with accent detection requires a modeling method that can capture speech pattern changes. Here, we study the accentual influences on Malay speakers' pronunciation and its prosodic impacts towards ASR system for CA language with seven Quranic accents identification. The proposed ASR system was developed over three stages. First, a dataset of *Surah Al-Fatihah* recitation was recorded from 14 Malay speakers in seven Quranic accents, forming a total of 5,684 words. Second, various spectral and prosodic features are extracted from the dataset for further classification process. The final stage includes training and testing the classification model. The existing ASR systems are often enabled by Gaussian Mixture Models (GMM) because of its capability to represent a wide range of sample distributions. However, GMM is susceptible to overfitting when the model complexity is high, due to the presence of singularities. To support identification of seven Quranic accents, Universal Background Model (UBM) is adapted to GMM using Maximum A Posteriori (MAP) estimation method. The UBM models were trained over each of Quranic accents, and combined to establish final UBM with 512 mixture components. The proposed ASR system utilizing the GMM-UBM outperformed k-NN, GMM, and GMM-iVector in identifying *Al-Fatihah* recitation to the corresponding Quranic accents. The GMM-UBM yields a testing accuracy of 86.148%, which is an increment of 4.435% from utilizing GMM alone.

INDEX TERMS Automatic speech recognition (ASR), Gaussian mixture model-universal background model (GMM-UBM), Malay speakers, Quranic accents.

I. INTRODUCTION

Accent is a pattern of pronunciation that distinguishes a person's speech as belonging to a certain linguistic group. The accent can cause variability in speech patterns and is unique to a specific individual, geographical areas, or ethnic cultures [1]. Automatic Speech Recognition (ASR) system is a human-computer interface that provides speech-recognition services for a wide range of applications. However, the development of the ASR system can be challenging due to variability in speech patterns among individuals [2]. The speech patterns can differ in terms of audio quality, pronunciation, vowels and consonants distinction, stress, and prosody. The variability of these components can reduce the recognition accuracy and consequently impair the overall performance of the ASR system [3].

In Arabic language, the variability of speech patterns is concentrated on accents in Classical Arabic (CA) language or also known as Quranic accents. The Quranic accents are styles used during reciting the Quran. There are 20 Quranic

The associate editor coordinating the review of this manuscript and approving it for publication was Mounim A. El Yacoubi.

accents in total, but only seven Quranic accents will be focused on this research, known as *Hafs, Khalad, Khallaf*-facet 1, *Khallaf*-facet 2, *Bazzi, Qunbul,*and *Ruwais*. These Quranic accents are permitted and acceptable in Islam [4], [5]. There is no conflict between one accent with another, since all the Quranic accents are revealed by Allah SWT and derived from the narration that came and linked to the Prophet Muhammad (PBUH) [6]. The revelation of the Quranic accents allows for adaptability when reciting the Quran regardless of geographical backgrounds and regions of the Muslims speakers.

However, there is a significant phonetic difference between Arabic and Malay languages. The Arabic language has a high lexical stress system compared to any other languages [7], including the Malay language. The emphatic and pharyngeal phonemes in the Arabic language are pronounced with lexical stress and pitch accent [8], [9]. In contrast to the Malay language, most phonemes are not stressed at all, with just a few phonemes being slightly stressed [9], [10]. This causing mispronunciation of articulation point (*makhraj*) of the CA phonemes [11], [12], as well as misarticulated Arabic phonemes [8], [13], [14], [15] (see Table 1) among the native Malay speakers when reciting the Quran. The mispronunciation also contributed by the confusion with the Malay colloquial dialect [12], [14], [16]. For example, the CA phonemes of *Sad* (ص), *Seen* (س) and *Zain* (ز) are taken from the word /sˤi ra: tˤa/ (صِرَاطَ), /si ra: tˤa/ (سِرَاطَ) and /zi ra: tˤa/ (زِرَاطَ), respectively. Each word from those three words is found within the verse 6 and 7 of *Surah Al-Fatihah*, and originally taken from different Quranic accents, as listed in Table 4-(c) and Table 4-(d). These three phonemes are often confused by the Malay speakers and tend misarticulate by stressing out the phoneme of *Seen* (س) bolder and heavier compared to the phoneme of *Sad* (ص) and *Zain* (ز). In fact, the phoneme of *Zain*(ز) is a stress letter and should be pronounced in stress sound, while the phoneme of *Sad* (ص) should be pronounced in bold (heavy) sound, and the phoneme of *Seen* (س) need to be pronounced in thin (light) sound [7], [17], [18]. All of these individual characteristics of phonemes are referred as prosodic information.

Features of the prosodic information in the form of pitch, energy, duration, and spectral-tilt are critical in providing important information to recognize accents of CA phonemes in the Quran [19]. Nevertheless, the prosodic features are disregard in the previous ASR research for CA language [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], which led to reduced performance of the ASR systems in these studies. One of the main reasons prosodic features are disregard in the previous studies is the absence of the publicly accessible Quranic accents dataset. Establishing a database of Quranic accents is a challenging task since it requires trained and knowledgeable reciters in the Quranic accents. The absence of the public database of the Quranic accents remains an issue to the present day and a major obstacle in developing ASR system for CA language.

In the previous ASR research for Arabic language, the phonological information is extracted from the speech waveform via the spectral features only, using the conventional Gaussian Mixture Models (GMM) technique [31], [32]. For CA language, the early studies [20], [33] incorporated spectral features of Mel-Frequency Cepstral Coefficients (MFCC) and Hidden Markov Model (HMM) classification, using CMU Sphinx tool and Matlab. These studies only considered a single accent of *Hafs* in their ASR research. Later, the classification of the ASR system for CA language was improved by replacing HMM with GMM based modeling technique [21], [22], [23]. Using the similar MFCC feature and GMM classification, [28], [29], [30] presented the ASR research method for multiple Quranic accents. However, the performance of their ASR system remain limited for recognizing the multiple Quranic accents. This is because their ASR system depends entirely on spectral based-features for feature extraction and employed conventional classification technique, that supports only binary classification. Due to accents in CA language, the implementation of conventional classification such as GMM can results in overfitting, because of the existence of singularities, when the complexity of the model is high.

Therefore, suitable features, modeling strategy and classification technique are required in the development of the ASR system for CA language. To ensure that issues of overfitting and dealing with large training data can be addressed, we proposed an ASR methodology for CA language. The proposed methodology aims to allow accurate Quranic accents recognition by incorporating prosodic features (pitch, energy, duration, spectral-tilt) with spectral features (MFCC) and Gaussian Mixture Models - Universal Background Model (GMM-UBM) based modeling technique. In conventional GMM, standard Maximum Likelihood (ML) approach is utilized in training both target and non-target models. The GMM-UBM utilizes UBM to train models efficiently, known as adaptation [34]. The adaptation can enable training model of Quranic recitation by native Malay speakers adapted to the respective Quranic accents. Thus, increase the performance and robustness of the ASR system. To the best of our knowledge, GMM-UBM has not yet been incorporated in any previous ASR research for CA language that supports multiple Quranic accents recognition. The ASR system for CA language can be used as a self-learning tool for Quranic recitation to aid Muslims in reciting the Quran based on different styles of Quranic accents. The self-learning tool can potentially be developed as a web-based application or a standalone application either for desktop or mobile.

In this research, we prefer to implement machine learning that uses GMM-UBM classification rather than applying the deep learning method. It is because, we have already identified the suitable features to be extracted based on experts' experiences and knowledgeable Quranic expert. In this case, the accent information carried by prosodic features is crucial, due to improve the modeling and recognition of Quranic

accents. Here, GMM-UBM classification is used to validate the targeted and tested features, whether it is correct or incorrect, as well as to understand and interpret the model created. Contrast with the deep learning, where this method automates the feature extraction process and directly taken the data as input. Means, deep learning is used to eliminate some data of pre-processing that is typically involved with machine learning. By using this technique, all the machines from the machine learning technologies are in charge of performing the feature extraction and modeling process. As the result, the process of interpreting the model and decision making has become more challenging and difficult to execute, because of the requirement of the large amount of data and powerful computing resources [35]. Moreover, it is extremely expensive to train, where it requires high performance of GPUs and hundreds of machines (complex data models), which definitely increases the cost to the users.

Our interest in this research is to observe the accentual influences on the pronunciations of native Malay speakers, and its prosodic effects towards recognition of the seven Quranic accents namely *Hafs, Khalad, Khallaf*-facet 1, *Khallaf*-facet 2, *Bazzi, Qunbul,*and *Ruwais*. The recognition of the Quranic accents will be focused on the recitation of the first chapter of the Quran, *Surah Al-Fatihah*. The *Surah Al-Fatihah* is chosen in this study because it is mandatory to be recited by Muslims during prayers. The experiment started with a thorough evaluation of phonetic properties of a set of letters and phonemes, which pronunciation is almost similar in different accents. Prior knowledge about the accents provides valuable information for speaker profiling, which can be incorporated into the decision parameter and techniques to improve the system performance and efficiency. Additionally, the level of fluency based on the proper pronunciation of the CA phonemes among native Malay speakers is also analyzed, in order to address the misarticulated phonemes issue (see Table 1). During front-end processing, the spectral and prosodic features are extracted and both types of features are computed and implemented as suggested by [36] and [37]. A distinct feature representing each CA phoneme can help in identifying and distinguishing one accent from another. Initial acoustic features of the confused phonemes are selected by studying speech production. The classification technique of GMM-UBM is proposed to be implemented in this study, which is believed to be an ideal and effective technique in identifying the misarticulated CA phonemes among Malay speakers, discriminating the accents of the language, as well as to improve the recognition results that are involved with accent element. The performance results of GMM-UBM are measured and compared with other previous research (using GMM, k-NN, and GMM-iVector), based on the accuracy (Acc) and Equal Error Rate (EER), with regard to Quranic accents.

The contributions of this paper are summarized as follows:

- A dataset containing recitation of *Surah Al-Fatihah* in seven Quranic accents (novel aspect of research work) is developed to solve the present problem of lack of datasets in this field. The dataset is composed of training and testing sets.
- The proposed ASR system considers prosodic features (pitch, energy, duration, spectral-tilt) and spectral features as input to GMM-UBM. The prosodic features carry traits of accents, but these features are disregard in the previous ASR research for CA language, due to absence of Quranic accents dataset.
- The proposed ASR system includes the GMM-UBM technique to support the adaptation of multi-classification, thus, enables recognition of various Quranic accents. In contrast, previous ASR systems for CA language relied on conventional classification technique that supported only binary classification, while the use of GMM-UBM is being less common.
- To verify our findings, the classification performance of the GMM-UBM is compared with GMM, k-NN, and GMM-iVector using the same data samples of Quranic accents.
- The implementation of the proposed ASR system with multiple features (spectral features (MFCC) and prosodic features) and GMM-UBM classification technique have resulted in a robust Quranic accents recognition. The proposed methodology recorded an accurate recognition of 86.15% (test-set) and 90.26% (train-set) and surpasses other techniques. This result performance had shown a novel aspect of our method, which indicated the successful implementation of GMM-UBM, with a combination of prosodic and spectral features.

The outline of the paper is organized as follows. Section II reviews relevant literary works about Quranic accents and its background, the scenario of Quranic learning in Malaysia, and previous ASR research for CA language. Section III presents the method and classification procedure, which highlighted the process of the proposed methodology. Section IV then presents an experimental study and evaluation results, based on the performance comparisons with other approaches. Finally, the conclusions and future works are drawn in Section V.

II. LITERATURE REVIEW

A. THE HISTORY OF QURANIC ACCENTS

The Quran is the holy book for Muslims, which written and recited in CA language. The CA language is an ancient Arabic language used when the first verse was revealed [38], [39]. The CA language differs from Modern Standard Arabic (MSA) and Dialectal Arabic (DA). The MSA and DA languages tend to change over time and often adaptable to other dialects. In contrast, the CA language is safeguarded and secured against modification, as the recitation is being taught and passed orally from one person to another, to preserve the correct manner of articulations. Reciting the Quran properly is essential in Islamic practices, such as prayers. Committing deliberate errors while reciting the Quran is prohibited in Islam.

The variability of pronunciation for certain CA phonemes is particularly referred to as accents used for Quranic recitation, known as Quranic accents (known as *Qira'at* in Arabic). According to Ibn al-Jazari (d. 602/1206), the science of Quranic accents is a discipline of knowledge that studies the method of pronunciation or articulation of a word in the Quran, where every dispute occurs is referred to its narrators [40]. Based on the Prophet Muhammad (PBUH) words (*hadith*), and Muslim scholars upon the commentary of Ibn al-Jazari, there are ten modes of Quranic recitation, which are permissible by Allah to Muslims. Seven Quranic accents are from *Shatebiah's* way, while three Quranic accents are from *Al-Durrah* way [41]. Every single mode of Quranic accent or style of recitation is derived its name from the reciter itself, as a prominent scholar (*Qari*). Each reciter recited to two narrators (*rawi*), whose narrations represented for each Quranic accent (*Qira'at*), with a sum of up to 20 Quranic accents. In this research, only seven Quranic accents of *Surah Al-Fatihah* have been selected for evaluation, known as *Hafs, Khalad, Khallaf*-facet 1, *Khallaf*-facet 2, *Bazzi, Qunbul,* and *Ruwais*.

B. SCENARIO OF QURANIC LEARNING IN MALAYSIA

In Malaysia, Muslims are mainly native Malay speakers, who are considered as non-Arabic speakers. The majority of phonemes from Malay language were derived from the English language. On the other hand, Arabic is a morphologically rich language as compared to English and Malay languages [42], [43]. The recitation of the Quran among Muslims worldwide is often based on a single accent (*Hafs*) of the Quran, including in Malaysia. However, learning the multi-accents (Quranic accents) of the Quran is considered mandatory and understandings it is essential, as different narrators narrate the verses differently [17], [18]. Fluency in Quranic reading, either in prayers (alone or in congregation) or outside the prayers, is directly associated with the particular recitation style of Quranic accents used. Awareness about Quranic accents is important to reach the Muslims (e.g., Malaysians) globally and to ensure that the misunderstanding among Muslims can be prevented [44], [45]. Furthermore, learning Quranic accents has a strong association with the knowledge of Islamic jurisprudence (*fiqh*) in sectarian diversity and the knowledge of Quranic exegesis, which are subjected to multiple interpretations.

Research on ASR involving Quranic accents is still unclear and limited. Focusing on Malay speakers, the critical issues in the Quranic recitation are concentrated on the misarticulated phonemes, which are influenced by the Malay colloquial dialect and native language. The frequent misarticulated phonemes have close articulation sounds to each other according to the pronunciation in Malay language, although in fact, the articulation points are distinguished in CA language [12], [13], [14]. Moreover, the CA language is known to have lexical stress [8], [9], and pitch accented on designated consonants [7], which are referred to the prosodic elements that determine the accent. In Malay language, the sounds of certain phonemes are only slightly stressed, while the huge remainder are not stressed at all [10]. Even the location of stress in Malay syllables often vary and fluctuate [46]. Based on this, the significant differences in lexical stress between the Malay and the CA languages used in the Quran are clearly noticeable. Thus, it is common for Malay speakers to mispronounce the CA phonemes in the Quran, which composed of a large proportion of stressed words. Researchers and experts have identified Arabic phonemes that typically misarticulated by Malay speakers. The phonemes are frequently pronounced according to their mother language. This finding has been discovered after years of teaching Malay students from all levels of education [11], [12], [13], [15], [17], [18], (summarized in Table 1). The concerning issue related to the misarticulated CA phonemes had arises, because there is a huge gap between Arabic and Malay languages, based on the comparison of the phonological aspect. Here, all the listed misarticulated phonemes of Arabic consonants are identified as confused, mispronounced, and considered difficult to articulate properly among Malaysian students. Some phonemes of Arabic letters sound similar to each other when articulated according to Malay accent. In fact, these phonemes have different articulation points according to Arabic accents.

As presented in Table 1, all misarticulated phonemes are arranged into three different groups, where the cases of interaction (confusion) with other phoneme from different group (in same row) had occurred and have been reported. For example, the Malay students are confused and often misarticulated the phoneme of *Sad* (ص) as the phonemes of *Seen* (س), and often confused with the phoneme of *Theh* (ث), and so on. Both CA phonemes of *Sad* (ص) and *Seen* (س), need to be clearly identified because these phonemes correspond to different Quranic accents as highlighted in *Surah Al-Fatihah*-verse 6 and 7 (see Table 4-(c) and Table 4-(d)). In this case, *Sad* (ص) should be pronounced in bold (heavy) sound, and *Seen* (س) needs to be pronounced in thin (light) sound. Whereas, the phoneme of *Theh* (ث) has a closely articulated phoneme with *Seen* (س), but in fact the articulation point of *Theh* (ث) is totally differed. It is often articulated like a 'whistle' group (*Seen* (س), *Zain* (ز), and *Sad* (ص)), rather than their unique articulation point by mistake.

In [13], there are seven consonants of *Qaf* (ق), *Thah* (ظ), *Khah* (خ), *Ghain* (غ), *Hah* (ح), *Ain* (ع) and *Heh* (ه), whereby the sound from each letter have not successfully distinguished, as proved from the final results of the unacceptable values of formant frequencies that measured using spectrogram. Hence, these consonants are considered as difficult Arabic phonemes, which are commonly misarticulated by Malaysian primary school children. Meanwhile, students from Malaysian higher learning institutions are also reportedly making various pronunciation errors when reciting the Quran [8], [11]. The frequent mistakes that repetitively been made by most of the students are categorized as the errors of

TABLE 1. Misarticulated phonemes (Malay speakers).

Group 1			Group 2			Group 3		
CA Letters	Phonemes	IPA	CA Letters	Phonemes	IPA	CA Letters	Phonemes	IPA
ص	Sad	s^{ς}/S	س	Seen	s	ث	Theh	θ
ذ	Thal	ð/∂	ز	Zain	z	ظ	Thah	$ð^{\varsigma}$/∂
ت	Teh	t	ط	Tah	t^{ς}/T			
ص	Sad	s^{ς}/S	ز	Zain	z			
د	Dal	d	ذ	Thal	ð/∂			
س	Seen	s	ش	Sheen	ʃ			
أ	Aa	ʔ/E	ع	Ain	ʕ			
ك	Kaf	k	ق	Qaf	q			
ث	Theh	θ	س	Seen	s			
ت	Teh	t	ث	Theh	θ			
د	Dal	d	ض	Dad	d^{ς}/D			
ح	Hah	ħ/H	ه	Heh	h			
خ	Khah	x	غ	Ghain	γ			

CA articulation point, which involved the consonants *Teh*(ت), *Tah* (ط), *Dal* (د), *Ghain* (غ), *Hah* (ح), *Ain* (ع), *Heh* (ه), *Khah* (خ), *Hamza* (ء), *Jeem* (ج), *Sheen* (ش), *Yeh* (ي), *Dad* (ض), and *Theh* (ث), *Thal* (ذ), *Thah* (ظ). These mispronunciation errors might because of inability to grasp the techniques of reading the Quran appropriately. Whereas, Quran illiteracy is described as a lack of ability to be proficient at a certain level of Quranic recitation and decent understanding, which includes the correct practice of pronunciation rules (*Tajweed*), point of articulation (*makhraj*), attributes of letters (*sifat*), and smooth pronunciation of Quranic letters (*fasahah)*. In general, the results obtained from previous research, and observations made so far, show that the students' ability to learn the Quran, particularly according to the *Hafs* style of Quran, is still below the expectation [47], [48], [49].

C. RELATED WORKS

The CA language has received less attention from computational linguistics and modern speech researchers compared to MSA. Since the year 2000, only few studies focusing on the CA language [22], [23], [24], [25], [26], [27], and considered only Quranic accent (single accent), such as *Hafs*. The slow progress might because of the limited CA corpus available and the absence of a standard Quranic accents database, to be used as a reference in research and development activities. Due to this limitation, research on the Quran becomes an intriguing subject of exploration. In the earlier stage of ASR research in the Quran, a few studies have been explored and tested using the *Hafs* style of Quranic recitation. The research on the speaker-independent for Quran recognition system has started to be proposed by [33], which was developed using the Sphinx-IV framework. Here, the HMM classification has been used for acoustic model training, and this research is purposely to contribute toward the development of a commercial assessment system, known as HAFSS [50], [51]. However, the HAFSS system does not enable users to recite from any place of the Quranic chapter and does not accommodate the non-native Arabic speakers. Meanwhile, the research by [20] has incorporated the spectral features of MFCC with HMM classification, which then been tested against the small Quranic chapter of *Surah Al-Fatihah*. Later on, HMM was replaced with GMM based modeling technique, which performed significantly better for the classification of the Quranic ASR system [21], [22], [23]. According to [22] and [23], the different numbers of Gaussian mixtures are used, along with the different numbers of HMM states, in modeling the CA phonemes. The different numbers of Gaussian mixtures need to be estimated, in order to obtain the most optimal results for recognition. In the meantime, the regression class tree has been conducted by [27], to estimate the values of a linear transformation of the mean and variance parameters of a Gaussian mixture-HMM system. Here, the 30th chapter of the Quran (*Juz Amma*) has been used as a dataset for training and testing, where the samples of recitation have been collected from five prominent Quranic reciters (*Qari*). Later, [21] has conducted a research about learning of the Quran language by developing the model and classifier using GMM. Here, a database also has been developed with randomized Quranic chapters, purposely to avoid the chapter bias and focus only on the reciters.

Previously, the ASR research work related to the variability of reading styles in Quranic accents was still ambiguous and less executed. It is due to the lack of research and progress that investigates on the pronunciation rules (*Tajweed*) error identification toward the Quranic accents recitations. Furthermore, the recognition of words is something that needs more improvement since the different accents of speakers need to be considered. Although the research related to Quranic accents was still less explored by researchers, but recently, a few research studies involving the multi-accents of the CA language (Quranic accents) have started to gain momentum, where a few researchers have been pioneer in this field. Here, their main research are focused only on the spectrogram voice analysis [52] and extra vowel analysis [41], but the status of classification are unknown, due to unclear discussion from both research. Later studies of Quranic accents [28], [29], [30], used a similar type of dataset like the previous researchers where the audio samples of Quranic accents were downloaded from the internet, which considered as unclean of speech samples. The audio samples are contaminated with unwanted elements such as noises, echo, and reverberation effects. These elements need to be eliminated as they might influence the overall performance results. Here, their research concentrated on echo cancellation technique during front-end feature extraction. Then, the conventional classification techniques of GMM, k-Nearest Neighbor (k-NN), and Probabilities Principal Component Analysis (PPCA) are implemented, where a combination of Affine Projection (AP) (front-end) and GMM achieved a higher accuracy rate compared to other techniques.

In general, the above-mentioned research (Quranic accents) and most past ASR research for Arabic and CA languages used in the Quran (a single accent of *Hafs*) have been designed with the conventional classification technique of GMM, which was performed solely with binary classification. The GMM has a great risk to suffers from the problem of overfitting when the model complexity is high [53]. This problem occurred under the presence of singularities, which associated with the Maximum Likelihood (ML) framework applied to Gaussian Mixture models. During the fitting process of a GMM to a dataset, when the variance gets to zero, the likelihood function of the Gaussian component becomes infinity. Thus, the model becomes infinity (overfitted). Here, the zero value of variance happens when there is only one point that leads to a singular covariance matrix, which occurred in the multi-variate Gaussian case. Furthermore, training GMM is time consuming, due to the great amount of speech data that need to be processed, resulting in a significant number of mixture components [54]. In [55], the authors found that the GMM based ASR systems are also susceptible to feature variability caused by non-language factors, such as speakers and channel distortions.

Based on the previous report, the implementation of GMM in this study that involves the accent of speakers, and a variety of Quranic accents could results in poor performance. Therefore, we investigate alternative classification techniques to deal with these problems, such as k-NN, GMM-iVector, or GMM-UBM. K-NN is a non-parametric algorithm and a popular learning algorithm because of its simplicity. K-NN uses data from several classes to predict the classification of a new data point. However, k-NN will suffer a slow prediction stage (high computation time), if the dataset is large and the dimension of features is high [56]. Other than that, k-NN is also sensitive to noise in the dataset, and any irrelevant features such as missing values and outliers, need to be manually imputed and removed, respectively [57]. Contrarily, GMM-iVector can perform well in noisy environment, while suffer and deteriorate in quiet condition [58]. Moreover, GMM-iVector requires a massive development of data, which costs a lot in most cases [59], [60]. In the presence of utterance duration variability, the GMM-UBM outperforms the GMM-iVector for very short utterances [61], while GMM-iVector requires utterances length more than two minutes.

Despite various progress made in previous studies, a reliable ASR system with Quranic accents identification remains an open research issue. The constraint issue associated with accents and phonological patterns of CA language in Quranic accents needs to be taken into consideration by ASR researchers for enhancement. To support various Quranic accents identification in ASR system, a large dataset is required. Training a large amount of data may result in overfitting, which can be prevented if an adaptive strategy is used. The implementation of the binary classification of the conventional classification approach for multi-classification seems impractical and unsuitable. Therefore, an appropriate modeling strategy and a suitable model with an ideal classification technique need to be developed. It is vital to overcome the overfitting issue, handling a large amount of training data, as well as identifies the speaker's recitation to the corresponding Quranic accents. The above reasons have inspired this study to explore the use of GMM-UBM based modeling technique in developing the ASR system with Quranic accents identification.

In this study, we described a novel method for recognizing the Quranic accent-based ASR for non-Arabic speakers (Malay), according to the respective class of Quranic accents by using the GMM-UBM. The way this algorithm performed modeling and make final decisions of the classes of Quranic accents could be considered as an early approach and the best initiative made so far, especially for the ASR research of Quranic accents. The GMM-UBM is being less common to be implemented in ASR research for Arabic language (particularly CA for Quranic accents), as compared to the other conventional classification techniques, as previously discussed in the literature. Besides, the accentual influences and prosodic impacts on Malay speakers' pronunciation also been studied, due to identify the suitable features for recognition process of Quranic accents. The proposed method is designed in such a way, purposely to overcome the research gaps and problems from the previous research in the Arabic language, especially CA. The goal is to improve the Muslim's learning process by inventing a self-learning tool, as well as to contribute to the advancement of ASR research of the CA language, particularly Quranic accents.

III. METHODOLOGY

The detailed methodology of this research is described in this section. Generally, the proposed study consists of three primary stages, denoted as data preparation, front-end processing, and back-end processing stages. In Fig. 1, a block representation of the overall methodology process is presented based on the stages highlighted.

First, identification of the data from the samples and data preparation, based on the scope of research (see Table 2) have been reviewed in section (A) under experimental setup. Second, the execution of the front-end processing stage, which involved the prosodic features of pitch, energy, duration, and spectral-tilt, whereas the spectral features are represented by MFCC. Third, training and testing phases during the back-end processing stage, where the execution of the proposed classification of GMM-UBM occurred. Both the second and third stages are two main stages in this research, where the front-end processing involved the preprocessing and feature extraction process, whereas the back-end processing focuses more on the modeling and recognition process. These two main stages have been discussed elaborately, under the section (B) of the proposed methodology for Quranic accents recognition.

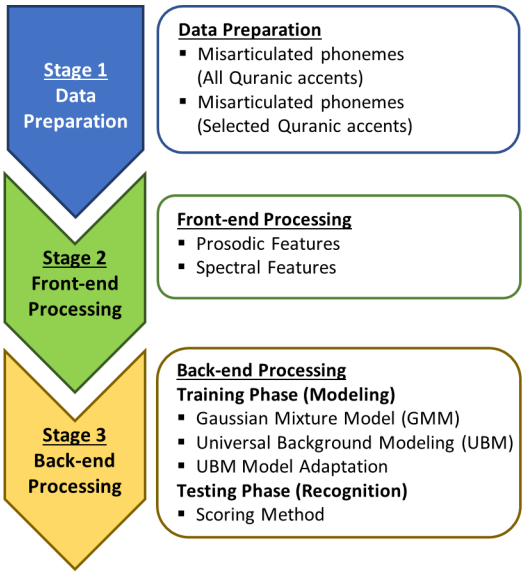

FIGURE 1. Overall methodology process.

A. EXPERIMENTAL SETUP

The experimental setup is prepared to conduct the experiment of the Quranic accents. As mentioned earlier, a lack of speech database, as well as the absence of a standard database for Quranic accents are identified as the major hurdles in conducting the ASR research of CA language in Quran. Due to this matter, we generate a local database by conducting in-house audio recordings, whereby the speech samples of non-Arabic speakers are compiled from the Malay reciters. Here, we collect the audio samples from 14 certified reciters (*Huffaz*), who have mastered the Quranic accents and obtained a diploma in *Qira'at* awarded by *Darul Quran, JAKIM*.[1]The purpose of collecting the data samples from these target speakers is because we want to ensure only appropriate Quranic accents are recorded for modeling (database); based on trusted speakers who had knowledge about Quranic accents. Moreover, before we carried out the assessment, the Quranic experts had already validated the audio samples. Meaning that, only clean and error-free audio samples are used for modeling and evaluation.

For this research, we selected the first Quranic chapter of *Surah Al-Fatihah* for evaluation. This chapter is mandatory to be recited by every Muslim while performing prayer (*Salah*); an act of Islamic worship that is considered as a pillar in Islam. Mistakes while reciting any verse in this chapter, especially during prayer, are prohibited in Islam, which results in invalidation of the prayer [45], [62]. Via ASR research, efforts in preserving this chapter have encouraged us to protect its meaning from distortion. Due to this reason, this research has motivated us to explore the ideal solution and remedy for developing the ASR system for Quranic accents. This research started with the recording process, where the audio samples of this chapter in seven different styles of Quranic accents were collected. During this process, speakers need to recite in a moderate tone, with proper recitation and correct articulation of the phonemes, according to the rules of pronunciation (*Tajweed*). Table 2 shows the setting used for the audio recording of Quranic accents, which is also considered as scope of research in preparing the datasets.

[1]*Darul-Quran*, JAKIM - Quranic Learning and Memorizing Institute, under Department of Islamic Development, Malaysia.

TABLE 2. Setting of audio recording for Quranic accents.

Items	Descriptions
Speech samples	• 14 non-Arabic speakers (Malay speakers) • Certified reciters - has a diploma in *Qira'at* from *Darul Quran, JAKIM, Selangor, Malaysia.* • 10 speakers - 5 females and 5 males (training) • 4 speakers - 2 males and 2 females (testing)
Quranic accents	• Tested on chapter *Al-Fatihah* in 7 Quranic accents (*Hafs, Khalad, Khallaf facet 1 & 2, Bazzi, Qunbul and Ruwais*)
Recording environment	• Quiet room (controlled environment)
Recording tool	• Digital voice recorder (OLYMPUS WS650S) • Audacity – sound editor software
Audio parameter	• Raw data (.wav) format • 16 bits and 16kHz sampling rate
Validation	• Audio samples validated by certified Quranic experts (*Huffaz*)

1) DATA PREPARATION

Each speaker needs to recite the complete verse of *Surah Al-Fatihah* in seven Quranic accents twice. Then, the recitation is paused in-between sentences, to generate the audio data with an approximate duration of 12 - 15 minutes per speaker. Overall, we successfully composed two main datasets, as shown in Table 3.

TABLE 3. Non-native Arabic speakers (Malay) datasets.

Dataset		Samples data – per Qira'at
Training	Modeling	• 5 males & 5 females (twice repetition) • 29 words x 10 speakers x 2 = 580 words • 580 words x 7 Qira'at = 4,060 words • 72% from overall samples
	Validation/ Development (train-set)	• 10-20% from 72% (modeling samples) used for train-set
Testing	Testing (test-set)	• 2 males & 2 females (twice repetition) • 29 words x 4 speakers x 2 = 232 words • 232 words x 7 Qira'at = 1,624 words • 28% from overall samples

The chapter of *Al-Fatihah* consists of eight verses/sentences, which is equivalent to 29 words per chapter. Each speaker needs to repeat the chapter twice, using seven Quranic accents. Hence, the database is composed of 1,568 verses/sentences, which is equivalent to 5,684 words (overall samples). The database is divide into training and testing sets, where the training set contains 72% of the data samples (4,060 words) and further divided into modeling set and development/validation set. Here, the percentage of development/validation set is allocated within the range of 10-20% from the 72% of modeling set (see Table 3). This dataset

is used to tune the hyper-parameters of the model during training, and to track the performance of the model created. The number of components for GMM (including UBM) is a hyper-parameter that requires to be set, relying on the quantity of data available. Thus, the number of components will vary between 16, 32 and 64, to determine the ideal setting for this approach. Meanwhile, the remainder 28% of the data samples (1,624 words) is used for testing. This data distribution was intended to avoid speaker and session bias, where data samples from different speakers were assigned to testing and training sets. In all cases, the data are gender-balanced. Data samples for each accent are equally distributed, with 232 samples/words (testing) are assigned to each accent. For running the experiment, we use MATLAB programming software [63]. The rules of recitation for each Quranic accent is depend on the different cues of phonemes, which occur for each verse in every Quranic chapter. In other words, each Quranic accent must be recited differently, using different phonemes at the places where the accent took place.

In this research, data limitation is not an issue, since our proposed dataset was obtained from 14 speakers (see Table 3), which is considered enough and sufficient for acoustic and phonetic analysis as indicated by [9] and [64]. In fact, this dataset is larger than the dataset used by [29], [41], and [65] for dialect analysis. The data augmentation technique is clearly restricted and difficult to be applied in this research, due to Islamic rules that permitted Muslim to recite only the correct of Quranic recitation based on pronunciation rules (*Tajweed*) guideline. Thus, any changes of audio data sample that clearly change the manner of pronunciation (against the *Tajweed* rules) are forbidden in Islam. Based on knowledge of Quranic accents (*Qira'at*), only certain CA letters and their phonemes were allowed to be recited differently, where each CA letter represented a different type of Quranic accents. Thus, observation must be made by identifying the location of the CA letters, which are labeled as Quranic accents phonemes within each verse. Also, we need to verify whether the Quranic accents letters have been listed under the misarticulated phonemes or not. We must be aware of any common mistakes that the Malay speakers have often made previously, due to avoid and minimalist the mispronunciation errors. In this study, we categorized each of verse in *Surah Al-Fatihah* into two different categories (sub-sections (a) and (b)), based on the phonological variations of the Quranic accents and misarticulated phonemes, as described below:

a: MISARTICULATED PHONEMES (ALL QURANIC ACCENTS)

There are several phonemes that are identified as misarticulated phonemes,as highlighted in blue (see Table 4-(a)). These phonemes are described as closely articulated phonemes, which are commonly mispronounced among Malay speakers (see Table 1). Here, the CA phonemes in these verses are not related to any differentiation of Quranic accents. Hence, any mistakes made while pronouncing the CA phonemes in these verses are prohibited in Islam.

b: MISARTICULATED PHONEMES (SELECTED QURANIC ACCENTS)

The same verse of the Quranic chapter might have a variety of Quranic accents. It depends on the verse's location in the Quranic chapter, which is narrated differently by the previous Islamic scholars. In this sub-section, the CA phonemes that classified as misarticulated phonemes have occurred at the places where the Quranic accents have taken place. Here, those phonemes are placed in certain words within the verse, which present the differences based on the Quranic accents (underlined and marked in red), as shown in Table 4. Those letters are located under specific words in verses 4, 6, 7-part 1, and 7-part 2 in *Surah Al-Fatihah*, which are classified based on different Quranic accents. In this case, any changes that occurred within the recitation, in regard to the Quranic accents is acceptable in Islam (pronounced differently using the certain phonemes of Quranic accents). Otherwise, if the articulation is mistakenly recited with a wrong phoneme, it is clearly forbidden in Islam.

TABLE 4. (a) *Surah Al-Fatihah* (verse 1,2,3, and 5) – all Quranic accents.

(a)

Verse number	Verse	Phonemic transcription
Verse 1	بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ	/bis mil/ - /la ħir/ - /raħ ma: nir/ - /ra ħi:::m/
Verse 2	ٱلْحَمْدُ لِلَّهِ رَبِّ ٱلْعَٰلَمِينَ	/ʔal ħam du/ - /lil la hi/ - /rab bil/ - /ʕa: la mi:::n/
Verse 3	ٱلرَّحْمَٰنِ ٱلرَّحِيمِ	/ʔar raħ ma: nir/ - /ra ħi:::m/
Verse 5	إِيَّاكَ نَعْبُدُ وَإِيَّاكَ نَسْتَعِينُ	/ʔij ja: ka/ - /naʕ bu du/ - /wa ʔij ja: ka/ - /nas ta ʕi:::n/

Based on Table 4-(c) and Table 4-(d), the CA phonemes that are categorized as misarticulated phonemes occurred at the places where the accent takes place within the verse. In this case, the CA phonemes involved are *Sad* (ص), *Zain* (ز) and *Seen* (س) from words /sˤi ra: tˤal/, /zi ra: tˤal/ and /si ra: tˤal/. The prosodic phonological features, such as word stress, pitch, duration, and speech intensity are evaluated for each phoneme using the Praat tool [66], due to prove theoretical concept as described in Table 5. Here, the phoneme of *Sad* (ص) does not exist in Malay language (see Table 5), and often confused with phonemes of *Seen* (س). Thus, Malay speakers often mispronounced both phonemes, due to the Malay colloquial dialect and differences in mother language.

The analysis of intensity show that, the phoneme of *Zain* (ز) has the highest energy (intensity) value, while the phoneme *Seen* (س) has the lowest energy (intensity) value.

TABLE 4. *(Continued.)* **(b)** ***Surah Al-Fatihah*** **(verse 4). (c)** ***Surah Al-Fatihah*** **(verse 6). (d)** ***Surah Al-Fatihah*** **(verse 7, part 1). (e)** ***Surah Al-Fatihah*** **(verse 7, part 2).**

(b)

Quranic accents	Verse	Phonemic transcription
• 'Aasim (Hafs) • Ya'akob (Ruwais)	مَالِكِ يَوْمِ ٱلدِّينِ	/ma: li ki/ - /jaw mid/ - /di:::n/
• Hamzah (Khalad) • Hamzah (Khallaf) – facet 1 • Hamzah (Khallaf) – facet 2 • Ibn Kathir (Bazzi) • Ibn Kathir (Qunbul)	مَلِكِ يَوْمِ ٱلدِّينِ	/ma li ki/ - /jaw mid/ - /di:::n/

(c)

Quranic accents	Letters	Verse	Phonemic transcription
• 'Aasim (Hafs) • Ya'akob (Bazzi)	(ص)	ٱهْدِنَا ٱلصِّرَٰطَ ٱلْمُسْتَقِيمَ	/ʔih di nasˤ/ - /sˤi ra: tˤal/ - /mus ta qi:::m/
• Hamzah (Khalad) • Hamzah (Khallaf) – facet 1	(ز)	ٱهْدِنَا ٱلزِّرَٰطَ ٱلْمُسْتَقِيمَ	/ʔih di naz/ - /zi ra: tˤal/ - /mus ta qi:::m/
• Hamzah (Khallaf) – facet 2 • Ibn Kathir (Qunbul) • Ibn Kathir (Ruwais)	(س)	ٱهْدِنَا ٱلسِّرَٰطَ ٱلْمُسْتَقِيمَ	/ʔih di nas/ - /si ra: tˤal/ - /mus ta qi:::m/

(d)

Quranic accents	Letters	Verse	Phonemic transcription
• 'Aasim (Hafs) • Ya'akob (Bazzi)	(ص)	صِرَٰطَ ٱلَّذِينَ أَنْعَمْتَ عَلَيْهِمْ	/sˤi ra: tˤal/ - /la ðiː na/ - /ʔan ʕam ta/ - /ʕa laj him/
• Hamzah (Khalad)	(ص)	صِرَٰطَ ٱلَّذِينَ أَنْعَمْتَ عَلَيْهُمْ	/sˤi ra: tˤal/ - /la ðiː na/ - /ʔan ʕam ta/ - /ʕa laj hum/
• Hamzah (Khallaf) – facet 1	(ز)	زِرَاطَ ٱلَّذِينَ أَنْعَمْتَ عَلَيْهُمْ	/zi ra: tˤal/ - /la ðiː na/ - /ʔan ʕam ta/ - /ʕa laj hum/
• Ibn Kathir (Ruwais)	(س)	سِرَاطَ ٱلَّذِينَ أَنْعَمْتَ عَلَيْهُمْ	/si ra: tˤal/ - /la ðiː na/ - /ʔan ʕam ta/ - /ʕa laj hum/
• Hamzah (Khallaf) – facet 2 • Ibn Kathir (Qunbul)	(س)	سِرَاطَ ٱلَّذِينَ أَنْعَمْتَ عَلَيْهِمُو	/si ra: tˤal/ - /la ðiː na/ - /ʔan ʕam ta/ - /ʕa laj hi muː/

(e)

Quranic accents	Verse	Phonemic transcription
• 'Aasim (Hafs)	غَيْرِ ٱلْمَغْضُوبِ عَلَيْهِمْ وَلَا ٱلضَّآلِّينَ	/ɣaj ril/ - /maɣ dˤu: bi/ - /ʕa laj him/ - /wa laːdˤ/ - /dˤaː:::l liː:::n/
• Ya'akob (Khalad) • Hamzah (Khallaf) – facet 1 • Hamzah (Ruwais)	غَيْرِ ٱلْمَغْضُوبِ عَلَيْهُمْ وَلَا ٱلضَّآلِّينَ	/ɣaj ril/ - /maɣ dˤu: bi/ - /ʕa laj hum/ - /wa laːdˤ/ - /dˤaː:::l liː:::n/
• Hamzah (Khalaf) – facet 2 • Ibn Kathir (Bazzi) • Ibn Kathir (Qunbul)	غَيْرِ ٱلْمَغْضُوبِ عَلَيْهِمُو وَلَا ٱلضَّآلِّينَ	/ɣaj ril/ - /maɣ dˤu: bi/ - /ʕa laj hi muː/ - /wa laːdˤ/ - /dˤaː:::l liː:::n/

TABLE 5. **Features and properties of phonemes** ***Sad*** **(ص),** ***Seen*** **(س), and** ***Zain*** **(ز).**

Features	Phonemes		
	Sad **(ص)**	***Seen*** **(س)**	***Zain*** **(ز)**
Consonant	Fricative		
Phonation	• Unvoiced	• Unvoiced	• Voiced
Manner of articulation	• **Pharyngealization** (pharynx/epiglottis is constricted)	• Tighten air movement to produce a disturbance	
Place of articulation	Alveo-dental Alveolar consonants		
	• Heavy(*tafkhiim*) letter • Sticking features • Pronounce with bold sound	• Thin/light letter • Pronounce lightly	• Stressed letter • Pronounce-stress sound
Distinctive Phonetic Features (DPF)	• Only found in Arabic language (unique)	• Arabic, English, and most languages	• Arabic, English, and most languages

The energy (intensity) of the phoneme of *Sad* (ص) is within the mid-range value between those two phonemes. This finding has supported the study made by [7], where the phoneme of *Zain* (ز) should be pronounced with a stressed sound. Whereas, the phoneme of *Seen* (س) needs to be pronounced with a thin (light) sound, while the phoneme of *Sad* (ص) should be pronounced with a bold (heavy) sound. Thus, the phonological variations of each phoneme need to be clearly identified, and the capability to detect the slight variations or differences between each CA phoneme depends on the robust Quranic accents recitation developed. Based on this, the prosodic features that carry the accent information are truly necessary for development. To see the phonological differences between these phonemes, we have measured the intensity values for word stress, as presented in Table 6 in Section IV.

As we mentioned earlier, the audio samples collected are considered clean and error free, which had been validated earlier by the experts. Therefore, we expected the mispronunciation errors should be none or minimal if any. In Section III-A(1), sufficient and adequate information extracted from the speech signal is essential and necessary for this research, due to cater the misarticulated phonemes issue. Whereas, the differences of particular phonemes at the places where the accent took place need to be clearly identified, in order to classify the related phonemes with targeted classes of Quranic accents. Therefore, the right choice of the optimal feature extraction algorithm and ideal classification technique is important, to distinguish the Quranic accents more accurately.

It is noteworthy to inform that, the improvement result as discussed later in Section IV concerns to all CA phonemes, even though in this Section III, we only concentrated on the phoneme of *Sad* (ص), *Zain* (ز) and *Seen* (س) from words /sˤi ra: tˤal/, /zi ra: tˤal/ and /si ra: tˤal/, as our examples. Our intention is to show the differences between

those phonemes by comparing their features and manner of articulations, in regard to different Quranic accents. Therefore, in the following section, we will discuss how the ideal classification process can detect the slight differences between each CA phoneme (i.e., *Sad, Zain,* and *Seen*), especially the misarticulated phonemes that occurred at the words of Quranic accents (highlighted in red in Table 4-(c) and Table 4-(d)).

B. PROPOSED METHODOLOGY OF QURANIC ACCENTS RECITATION RECOGNITION

In this section, the proposed method of Quranic accents recognition is described in detail. As presented in Fig. 2, the block diagram consists of two primary components: front-end processing, and back-end processing. For front-end processing, the acoustic waveforms are transformed into more compact and less redundant representations, known as 'acoustic features'. This process combines spectral (MFCC) and prosodic features. The back-end processing involved training (modeling) and testing (recognition) phases. Here, the GMM-UBM classification and scoring method will be implemented, respectively. However, only the proposed classification technique of GMM-UBM will be highlighted in this paper.

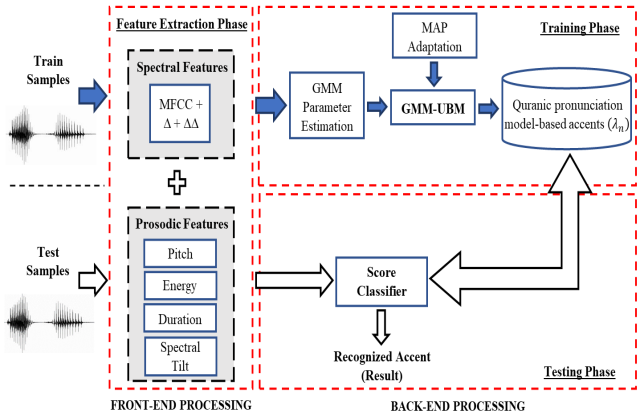

FIGURE 2. Block diagram of proposed method for Quranic accents recitation recognition.

1) FRONT-END PROCESSING

The previous ASR research in the CA language [22], [23], [24], [30], [67] had extracted the phonological information from the speech waveform through spectral features, making it difficult to recognize the specific characteristics and traits of an accent. Relying only on spectral features (e.g., MFCC) for feature extraction may impair the performance of ASR system in the presence of accents [3], [68]. Thus, using only spectral features in this study can be insufficient, since the CA language involves multiple Quranic accents. Variations of accents and speech patterns in Quranic accents will require additional features such as prosodic, to obtain adequate information for speech analysis in the Quranic recitation. Prosodic features correspond to the suprasegmental speech of the energy (stress), pitch, duration and spectral-tilt [19], which carries the attributes of accent in CA phonology. Therefore, we believed by extracting the spectral and prosodic features from the CA phonemes, sufficient information from each CA phoneme can be obtained, and thus, tackling the issue of misarticulated phonemes. In this research, we implemented both features of spectral and prosodic altogether, purposely for Quranic accents recognition.

a: PROSODIC FEATURES

Prosody deals with the acoustic qualities of a sound, which represent the prosodic information of the speech, such as rhythm, stress, intonation and so on. These prosodic features is important for the intelligibility and efficiency of Arabic as a stress-timed language [69]. The lexical stress is phonetically recognized through manipulation of four prosodic features variables, denoted as: (1) energy (intensity); (2) pitch (fundamental frequency (F0)); (3) duration; and (4) spectral-tilt. The consonants encapsulate to the acoustic features from the word accent are thoroughly analyzed, where significant variation of features helped in distinguishing one accent from another. It will concentrate on the multiple Quranic accents, which had been tested on Malay reciters.

In this study, pitch, energy, duration, and spectral-tilt measurements are extracted over the full waveform or longer speech segments (sentence, word, and syllable). The pitch (F0) is estimated by the fundamental frequency approximated by the pitch tracking algorithm, while the energy is computed by the root mean square value of each sample. The level of energy able to helps in identifying the voiced/unvoiced region of speech, together with pitch and duration to represent the stress pattern of speakers. Here, the duration represents the variation of one's speaking style length of spoken segment, which influences by speaker's accent and dialect. Meanwhile, the spectral-tilt values are calculated as the slope of the FFT (extracted over a window of 20-ms and shifted every 10-ms), where the use of this type of feature is inspired by the findings in [70].

b: SPECTRAL FEATURES

Features extracted from the vocal tract system are referred to as spectral system or segmental level of features. Fig. 3 shows a block diagram of spectral features for front-end processing, where a widely used feature extraction of MFCC was proposed to represent the spectral features in this ASR research.

(i) PRE-PROCESSING OF SPEECH SIGNAL: The sampling frequency of the recorded speech is set at 44.1 kHz. Before the preprocessing started, each sample is converted into .wav format and downsampled to 16 kHz manually, using the sound editor software called Audacity. In the early stage of

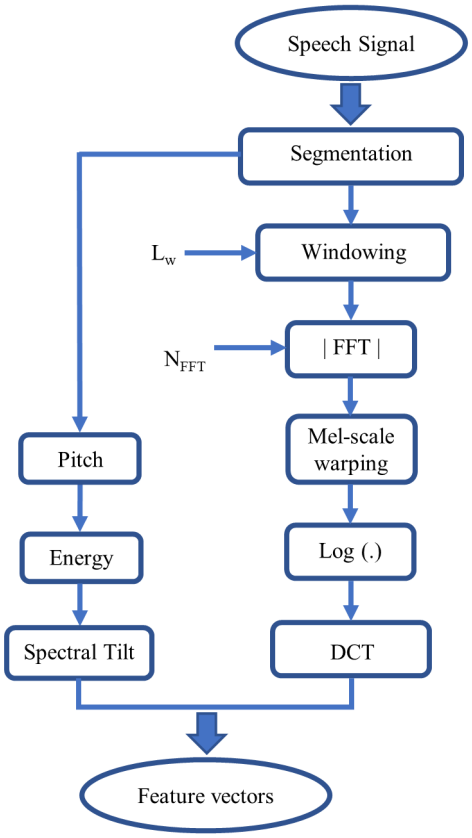

FIGURE 3. Block diagram of spectral features for front-end processing.

preprocessing, the speech samples, $x\,(n)$ is pre-emphasized using the first-order highpass filter. Then, we estimate the frame-level of spectral features (MFCC), as well as the frame-level of pitch from the filtered waveform. Based on [71], spectral features are used for modeling the variation of pitch pattern from the speaker's speech. Instead of using the average spectrum over multiple pitch cycles, the spectrum derived from each pitch cycle is also considered for feature extraction. Here, analyzing the spectral features of each pitch cycle, where each pitch period of a speech signal may provide more distinctive info related to spoken speech.

(ii) FEATURE EXTRACTION: After preprocessing of speech signal, the filtered speech samples $x_t\,(n)$ that containing the frame-level of pitch, and frame-level spectral features (MFCC) are subjected to parameterization process through degradation. This process is achieved by multiplying the following 25-ms Hamming window function to compensate for the overlapping between the neighboring frames and to minimize the signal discontinuities within the frames. Next, due to analyze the output $x_t\,(n)$ obtained in frequency domain, the N-point Fast Fourier Transform (FFT) of Discrete Fourier Transform (DFT) is first applied to compute the spectral coefficients from each frame and converted them into frequencies. Then, the result is further weighted by a series of triangular filters to gain Mel spectral coefficients. Last, the MFCC values (static features) are obtained by computing the logarithmic value of Mel-scale and performing the Discrete Cosine Transform (DCT) on the resulting spectrum, plus the energy term (the 13^{th} parameter). The additional process after the last, we computed the time derivatives of MFCC, to obtain information about the velocity (delta) and acceleration (double delta) of each feature vectors. The final features (static features) of MFCC do not capture the dynamic in the spectral changes. Thus, the time derivative is essential to get temporal information. By adding and expanding the time derivatives of delta and double delta to the basic static features of MFCC, the efficiency of ASR can be substantially improved [72].

Based on the final features of MFCC, the first 12 MFCC coefficients represent the full band of spectrum, of which the first coefficient (C_1) of MFCC is denoted as spectral-tilt. We used the information obtained from this spectral-tilt, because the lower order cepstral coefficients of MFCC contain most of the signal information about the overall spectral shape of the source-filter of transfer function. The lower order of MFCC cepstral coefficients also able to improve the performance of ASR [73].

2) BACK-END PROCESSING

In this section, we present two different stages of back-end processing, which are training and testing stages. For training stage (modeling), we next describe the specific components of the proposed classification of GMM-UBM, including its adaptation in developing the accent model. Meanwhile, for testing stage, the scoring method will be discussed for recognition process.

a: GAUSSIAN MIXTURE MODELS (GMM)

The GMM is a well-studied statistical method and unsupervised classification technique, which is used to model language and accent. It consists of a number of Gaussians to provide multi-modal density representation of each model. In pattern recognition, GMM has been used to generate speaker models with different accents, and to match the different patterns to that of the trained models. In this research, we used GMM to train and model the phonetic sound of multiple accents used for Quranic recitation by approximating the probability distribution. The first process of modeling of data is executed at the training stage. Our experiments operate on cepstral features, which are previously extracted from the front-end processing stage. Here, the feature vectors is represented by X, which is given by:

$$X = [x_1, x_2, x_3, \ldots, x_k, \ldots, x_T] \tag{1}$$

where k is the frame index, and x_k represent N dimensional MFCC from k^{th} frame, while T is the total number of features used to form feature vectors of X. These vectors are used to develop an accent model by training the Gaussian mixture models.

A Gaussian mixture density is a weighted of M component densities given by the equation:

$$p\left(x_k \left| \lambda \right.\right) = \sum_{i=1}^{M} w_i b_i \left(x_k\right) \tag{2}$$

where x_k is a N-dimensional feature vectors, component densities represent by $b_i\left(x_k\right), i = 1, \ldots, M$ and $w_i, i = 1, \ldots, M$ are the mixture weights.

Each component density is D-variate Gaussian function, given by following equation (3):

$$b_i\left(x_k\right) = 1/\left[(2\pi)^{D/2} \det\left(\sum_i\right)^{1/2}\right] \times exp\left\{-\left(1/2\right)\left(x_k - \mu_i\right)^S \left(\Sigma_i\right)^{-1}\left(x_k - \mu_i\right)\right\} \tag{3}$$

where, S is the transpose operation, μ_i is the mean vector and covariance matrix Σ_i. The mixture weights are normalized and satisfy the constraint of $\sum_{i=1}^{M} w_i = 1$.

The complete Gaussian mixture density is parameterized by mean vector (μ_i), covariance matrix (Σ_i) and mixture weights (w_i) from all component densities. These parameters are defined by the notation:

$$\lambda = \left\{w_i, \mu_i, \sum_i\right\}, i = 1, \ldots, M \tag{4}$$

In this research, we used these components densities to capture the information from the multiple Quranic accents. The number of components involved depends on the number of components used in the training, which might be different for each GMM state. Here, λ represents as a model for each accent of Quran. Means, each class of accent is given one GMM (i.e., λ). The primary goal of the training stage is to measure the best possible values of parameter for the λ, due to match the feature vectors distribution. The Maximum Likelihood Estimation (MLE) technique is applied to estimate the parameters of λ (equation (4)), which optimizes the likelihood of GMM for the training data. In this present work, to establish the model for multiple Quranic accents, we computed a set of observations by utilizing the feature vectors of X, as denoted from (1). The GMM likelihood can be written as:

$$p\left(X \left| \lambda \right.\right) = \prod_{k=1}^{T} p\left(x_k \left| \lambda \right.\right) \tag{5}$$

MLE is executed by applying the iterative procedure known as Expectation-Maximization (EM) algorithm. The EM model begins with a model λ and computes the new model $\widehat{\lambda}$, as denoted as $p\left(X \left| \lambda \right.\right) < p\left(X \left| \widehat{\lambda} \right.\right)$. The new model is considered the initial model for the following step, and the EM process is repeated until a convergence threshold is obtained. In this case, the EM iteratively fine-tunes the GMM parameters by increasing the likelihood value of the estimated model for the feature vectors observed.

b: UNIVERSAL BACKGROUND MODEL (UBM)

This section describes the form of the proposed GMM-UBM for Quranic accents recitation recognition. The modeling phase involved establishing a model that represents the phonetic or acoustic space of each speaker. This process is normally achieved with the help of statistical background modeling, from which the speaker-specific models are adapted. In order to model the multiple of accents in Quran successfully, we proposed to use the GMM-UBM classification technique at the training stage, which is described in Fig. 4.

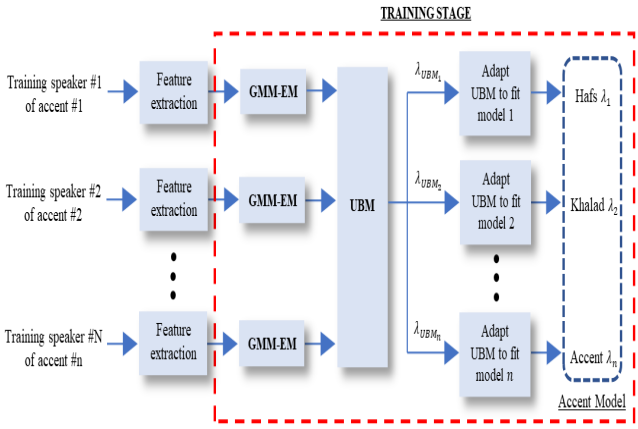

FIGURE 4. Architecture of the GMM-UBM training.

The UBM modeling technique is an enhancement to the GMM modeling technique. The UBM is a large GMM with the ability to model speaker-independent distribution. This capability makes GMM-UBM more suited to handle large training data than traditional GMM classification. Our intention in this research is to select the speech samples of Quranic recitation that reflect the expected alternate of Quranic accents data samples, to be encountered during recognition. Initially, the GMM-UBM method is used to select the train model, and also to determine the likelihood ratio for the testing speech sample along with the trained model and UBM.

Table 3 presents the speech data from 10 speakers per each Quranic accent in the database for training set. In the training, the data of all Quranic accents from 4 speakers were used to develop the UBM, and the remaining data of training set were used to adapt the UBM model. The process begin where the observation vectors (acoustic feature vectors) from the speakers' utterances are extracted (i.e., MFCC and prosodic features), for which a GMM statistical model is developed. Therefore, in the large observation set consisting of all the speakers, an acoustic model is established through EM algorithm for better converged of UBM. In this case, the acoustic model is created for the CA phonemes with its corresponding type of Quranic accent, as presented earlier in sub-section 2(a). By doing this, we estimate and create the universal background models of UBM, where its parameters form a baseline model for adaptive Maximum A Posteriori

(MAP) estimation methods. For this research, we trained the UBMs over each type of Quranic accents in the data (seven types of Quranic accents), and then combined (pooled) all Quranic accents models altogether, to create the final UBM with 512 mixture components. The pooled data for each Quranic accent are balanced over the male and female within the speech data. The GMM-UBM represented behavior and general characteristics of all the seven accents together after enrolment of speech samples from the enormous set of Quranic accents. For each target pattern (Quranic accents), a specific GMM will be gained by adapting the UBM through MAP criterion. We will clarify this step in the following sub-section 2(c).

As the same procedure executed in sub-section 2(a), every UBM model is a mixture that composed of M normal distributions which described as; mean vectors $\{\mu_1, \mu_2, \ldots, \mu_M\}$, $\{\Sigma_1, \Sigma_2, \ldots, \Sigma_M\}$ as full covariance matrices, and weighting factors $\{w_1, w_2, \ldots, w_M\}$, i.e., as stated in equation (2), (3) and (4). Later, seven Quranic accents models $\left(\lambda_{UBM_n}\right)$ are derived by adopting the UBM, using 4 speakers (approximately 28.8%) of training data.

c: UBM MODEL ADAPTATION

After calculating the Gaussian parameters of the UBM, the GMM for each class of Quranic accents is obtained. This adaptation process is performed using data from 6 speakers (approximately 43.2%). The class model of an accent in GMM-UBM system was derived by adapting the Gaussian parameters of the UBM using Bayesian adaptation. In the conventional classification of GMM (i.e., MLE), accent model is trained independently compared to UBM. Yet, in the adaptation method, the parameters of the accent models are derived by updating the trained parameters of UBM. The adaptation method offers advantages over the EM algorithm during the training of each accent model of GMM [74], [75], [76]. Through an adaptation, the model for a class is obtained by updating the parameters of the UBM, using the training data of the respective class. This method offers a tighter coupling between the class model of accent and UBM, which produces a much better performance compared to the decoupled models like traditional GMM. Where, the efficiency after performing the coupled approaches will not been affected by unseen acoustic event. Furthermore, all the accent models have similar initialization parameters, that are the same as UBM. On top of that, MAP adaptation integrates the robustly estimated of UBM parameters with the accent model parameters, which leads to more durable and robust estimation of accent models (accents with inadequate training data). Last, the training process of a new accent model in GMM-UBM has become faster than performing the EM algorithm in conventional GMM, which allows for a fast-scoring technique during the testing stage. Low computational complexity in GMM-UBM also can be considered as one of its attractive features, which suitable to be applied for any operation in real-time [54].

In adaptation-based model with GMM, we implemented the MAP estimation in the UBM adaptation-based model. The MAP algorithm involves two steps: (1) acquisition of information about the parameters to adapt the UBM for class estimation; and (2) the mixing of newly derived parameters with the old parameters and the models of UBM are updated using coefficients of data dependent mixing. This data dependent mixing is performed in such a way, where the mixture that are highly influenced by the accent specific data in the present class, able to keep maintains the parameters from UBM. In this part of the work, the mathematical computation involved the adaptation of UBM and class model of accent. Based on the articulations pertinent to a particular class of Quranic accent, the UBM model is subject to MAP adaptation for every speaker individually. Interpreting *a priori* probabilities is referred to the observation vectors set x_k belongs to the i_{th} acoustic class, as per described by equation (3). In this way, we expect to adapt the acoustic model to the certain group of speakers (i.e., Malay speakers) based solely on the CA phonemes of a particular class of Quranic accents.

Given the T training vectors from a class model of accent $X = [x_1, x_2, \ldots, x_T]$, for each mixture i in the UBM, we first identify the probabilistic alignment of the training vectors onto the Gaussian mixture components in UBM. For i_{th} mixture in the UBM, the relationship is mathematically described as follows:

$$P\left(i\,|x_t\right) = \frac{w_i P_i\left(x_t\right)}{\Sigma_{i=1}^{M} w_v P_i\left(x_t\right)} \tag{6}$$

where, w_i and w_v stand for the mixture of weights at the corresponding index. The $P\left(i\,|x_t\right)$ denotes the probability of frame x_t, given the mixture probability i, whereas M denotes the number of mixtures i. Lastly, $P\left(i\,|x_t\right)$ stands for the probability of mixture i, given the frame x_t. By utilizing this probability as stated from (6), the sufficient statistics and new parameters are computed as follows:

$$n_i = \sum_{i=1}^{T} P\left(i\,|x_t\right) \tag{7}$$

$$E_i\left(x\right) = \left(1/n_i\right)\left[\sum_{i=1}^{T} P\left(i\,|x_t\right) x_t\right] \tag{8}$$

$$E_i\left(x^2\right) = \left(1/n_i\right)\left[\sum_{i=1}^{T} P\left(i\,|x_t\right) x_t^2\right] \tag{9}$$

As shown in the above formula, we computed the new parameters $\left\{n_i, E_i\left(x\right), E_i\left(x^2\right)\right\}$ and sufficient statistics from the class model of accent developed for specific training data. T denotes a total number of frames, whereas n_i represents the posterior probability of the mixture i , and thus, it is called *count* moments. $E_i\left(x\right)$ represents the first-order moment, which indicates the expectation value of i_{th} mixture from the speech frames. $E_i\left(x^2\right)$ represents the second-order moment, describing the variance in the probabilities of i_{th} mixture from the speech frames. These new parameters used to update the

old UBM parameters (see sub-section 2(b)) for i_{th} mixture. By applying the adequate statistics, the adapted parameters for mixture i_{th} in the UBM are computed as follows:

$$\widehat{\omega}_i = \left[\frac{\alpha_i^{\omega} n_i}{T} + \left(1 - \alpha_i^{\omega}\right)\omega_i\right]\gamma \tag{10}$$

$$\widehat{\mu}_i = \alpha_i^m E_i\left(x\right) + \left(1 - \alpha_i^m\right)\mu_i \tag{11}$$

$$\widehat{\sigma}_i^2 = \alpha_i^v E_i\left(x^2\right) + \left(1 - \alpha_i^v\right)\left(\sigma_i^2 + \mu_i^2\right) - \widehat{\mu}_i^2 \tag{12}$$

The adaptation coefficients and parameters of $\left[\alpha_i^v, \alpha_i^m, \alpha_i^{\omega}\right]$ control the balance between the old and new parameters and sufficient statistics for the variances, means, and weights, respectively. For each parameter and each mixture, a data-dependent adaptation of coeffcient is defined as follows:

$$\alpha_i^{\rho} = \frac{n_i}{n_i + r^{\rho}} \tag{13}$$

where, r^{ρ} is a fixed relevance factor for parameter ρ. For example, in the language recognition system, r^{ρ} is considered as a number between 6 to 16. The use of parameter-dependent relevance factors allows for the tuning of different adaptation rates of the weights, means, and variances. This is to ensure that one of the experiments conducted, had analyzed the impact of different r^{ρ} on the overall performances of a system. The scale factor, γ, is computed over all adapted mixture weights to ensure the sum of roots of unity.

Based on the adaptation process described previously, the Gaussian parameters of the UBM from each utterance of phoneme *Sad* (ص), *Zain* (ز) and *Seen* (س) is adapted separately to the respective class model of Quranic accents, from which these phonemes are originated. Via adaptation, the right articulations of those three phonemes are able to be updated regularly. After the adaptation process is executed, seven Quranic accents models $\left(\lambda_n\right)$ are constructed, under the optimal training procedure.

d: SCORING METHOD

The testing stage evaluates if a test data is matched with the reference models (accent models), due to determine recognition performance. The recognition performance is measured on the remaining 28% of the overall dataset (see Table 3). Here, the pronunciation scores for an input is computed against all seven Quranic accents models. Then, the scores against all seven Quranic accents models of an input are further analyzed by the score classifier. The score classifier computes the average of log-likelihood to identify the highest score among the Quranic accents. The Quranic accent with the highest score is assigned as an output result for the input.

The evaluation process involved the use of a classifier belong to GMM-UBM, known as Scoring Method. For GMM-UBM evaluation, the adaptation of models' parameters from UBMs enables for a faster technique to evaluate the scores of the models. There are two steps involved in the evaluation process; first, we identified the top H scoring components in the UBM and the likelihood ratios in UBM are computed using only the top H components; second, the test vector is scored against the only corresponding H components in the adapted accent model to determine the likelihood of utterance [74], [76]. The estimations are done computationally using the pseudocode as follows:

For each frame $t = 1, 2, \ldots, T$
  For each component $k = 1, 2, \ldots, M$ compute

$$P_{ubm}\left(k, t\right) = w_k \times N\left(x_t \left|\mu_k, \Sigma_k\right.\right)$$

End

The $P_{ubm}\left(k\right)$ was sorted across t and the top H scores were selected, where $N\left(x_t \left|\mu_k, \Sigma_k\right.\right)$ is calculated as follows:

$$P\left(x_t \left|M_i\right.\right) = (1/T)\sum_{k=1}^{M} w_k \left(1/\left[(2\pi)^{D/2}\left|\Sigma_k\right|^{1/2}\right]\right) \times \exp\left\{-(1/2)\left(x_t - \mu_k\right)^S \Sigma_k^{-1}\left(x_t - \mu_k\right)\right\} \tag{14}$$

The T denotes the total number of frames, S represents the transpose procedure, M is the number of Gaussian components, D is the dimension of feature vectors and w_k is the weights of the components. The scoring method is illustrated in Fig. 5.

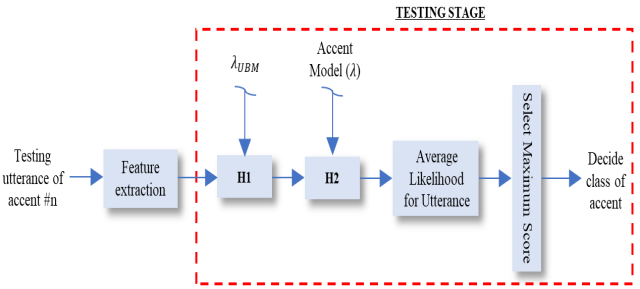

FIGURE 5. Overview of GMM-UBM Scoring method.

In the evaluation step, the classiffier returns the partial *score* (T, i), the probability value of the models trained for the $i - th$ evaluated class, where T is the input vector of the features obtained from the tested speech. The resulting class, I^* is derived from the maximum overall probability using the following equation:

$$I^* = arg \max_{1<i<N} score\left(T, i\right) \tag{15}$$

where, N is the number of all partial scores corresponding to the number of the classes.

In verification stage, we determine the score of features matrix of an unknown utterance X, for a speaker i, by the following equation:

$$\lambda\left(X, i\right) = (1/T)\left[\log p\left(X \left|\lambda_i\right.\right) - \log p\left(X \left|\lambda_{ubm}\right.\right)\right] \tag{16}$$

where, X consists of T number of frames. The scores obtained from the target and impostor trials are used to evaluate the system performance. Here, every phoneme of *Sad* (ص), *Zain*

(ز) and *Seen* (س) from the test-set sample is evaluated individually. These phonemes are verified based on the highest score value obtained, towards the respective class of Quranic accents as an output result. The percentage of precision is high if the output result of Quranic accent model is correct. In this case, the phoneme of *Sad* (ص) in word /sˤi ra: tˤal/, phoneme *Zain* (ز) in word /zi ra: tˤal/, and phoneme *Seen* (س) in word /si ra: tˤal/, should be verified as (*Hafs-Bazzi*), (*Khalad-Khallaf_1*) and (*Khallaf_2-Qunbul-Ruwais*), respectively (see Table 4-(c) and Table 4-(d)).

IV. EXPERIMENTAL EVALUATION

Assume that the acoustic features for each of accent are distinct, the acoustic models can use these distinctions to categorize the input data into groups. We constructed these models from a combination of each Quranic accent represented by spectral and prosodic features. The training samples from each accent are used to estimate the parameters of the model. The accent dependent models are further used to produce scores for their classification. In this section, the findings from the experiments are analyzed, to identify the best technique for CA language recognition with accent identification. The experiments were conducted on in-house database, where the data were recorded by the Arabic language experts. The results from both train-set and test-set are analyzed in this section, to identify the optimal features that can results in an accurate recognition. The results of test-set and train-set are reported separately, since the testing samples are obtained from two different sources of datasets, as mentioned earlier in Section III-A. In this research, a k-fold cross validation technique with 10-folds has been conducted, showing the average value of accuracy and EER for every experiment.

The recognition performance of seven Quranic accents used in the recitation of *Surah Al-Fatihah* are measured and presented in terms of *Accuracy* (Acc) and *Equal Error Rate* (EER). The following evaluation measurements are used to evaluate the performance, as described below:

$$Acc = 100 \times \left(\frac{Correct\ Acceptance + Correct\ Rejections}{Total\ number\ of\ samples} \right) \quad \text{or} \quad \frac{TP + TN}{P + N} \tag{17}$$

$$Err = \frac{FP + FN}{TP + TN + FN + FP} = \frac{FP + FN}{P + N} \tag{18}$$

The terminologies are defined as follows:

- True Positive (TP): Number of correctly recognized class samples.
- True Negative (TN): Number of correctly recognized samples that do not belong to the class.
- False Positive (FP): Samples that assigned incorrectly to the class.
- False Negative (FN): Samples that are not recognized as class samples.
- Positive (P): Number of real positive cases in the data
- Negative (N): Number of real negative cases in the data

A. FRONT-END PROCESSING

The performance of the prosodic features, spectral features, and the combination of both are evaluated and analyzed in this section. The assessment is performed to determine the most effective features for CA language recognition with Quranic accents identification.

1) PROSODIC FEATURES ANALYSIS

a: ACOUSTIC ANALYSIS

The phonetic information obtained from the acoustic analysis, which performed on the prosodic features explains the differences in the characteristics and manner of articulations (accent) between Malay and Arabic languages. The prosodic analysis of consonant phoneme can help define the synchronic differences between the Quranic accents. For this analysis, the acoustic and phonetic properties between the phonemes *Sad* (ص), *Seen* (س) and *Zain* (ز) will be differentiated, due to see how the prosodic features differ from one phoneme to another based on Quranic accents. Any differences in the characteristic and behavior of each phoneme on those three phonemes (*Surah Al-Fatihah*-verse 6) can be identified based on the different dominant properties of the phoneme-based accent. All samples and results were validated by the Quranic experts, based upon the CA syllables approach and theory that has been agreed by the Muslim scholars. The phonetic research focused on stress patterns in CA phonemes is performed, where the acoustic parameters, such as intensity, pitch and duration are extracted using speech analysis software Praat [66]. This tool is believed capable of measuring the prosodic and intensity values through spectrogram with a proper stress (*al-nabr*), while pronouncing the words of Quranic accents. The measured intensity values based on the word stress are presented in Table 6.

The results presented in Table 6 are generated from the audio samples collected from ten Malay speakers, the same speakers used as primary dataset for modeling stage. Based on the mean energy of intensity presented in the table, almost all respondents showed promising results; the finding indicated that all Malay speakers are capable of pronouncing and articulating the three words properly, within the acceptable values (prosodic results) as compared to the native Arabic speaker, as well as in reference to the fundamental theory of word stress in various Quranic accents [7], [11]. The results fit our initial expectation, since all ten respondents were drawn from the primary samples used in modeling stage, which mainly used for generating database of Quranic accents in this research. To develop a database with proper recitations in various Quranic accents, only audio samples from eligible Malay speakers who are experts and masters in Quranic recitation with Quranic accents are selected. With proper learning and training of CA phonemes, the mispronunciation issues while

TABLE 6. Energy (intensity) value.

		Energy (Intensity) value (dB)						
	Word	/sˤi ra: tˤa/ (صِرَاط)		/zi ra: tˤa/ (زِرَاط)		/si ra: tˤa/ (سِرَاط)		
	Syllable	/sˤi/ (ص)		/zi/ (ز)		/si/ (س)		
	Quran accents	Hafs	Bazzi	Khalad	Khallaf-facet 1	Khallaf-facet 2	Qunbul	Ruwais
Native Sp.	Median	59.2	58.3	68.9	67.2	58.4	54.6	58.2
	Mean energy	60.5	58.8	**68.8**	67.1	58.1	55.3	57.4
Male Speaker								
1	Median	51.6	52.4	54.4	54.4	52.3	53.2	50.7
	Mean energy	57.8	57.7	59.0	**59.0**	55.9	56.4	56.9
2	Median	48.4	55.3	56.6	58.3	48.4	45.9	47.4
	Mean energy	57.9	57.4	58.2	**61.1**	50.9	53.2	56.1
3	Median	58.1	63.3	67.0	65.3	52.6	49.2	52.8
	Mean energy	62.2	63.7	68.2	**69.2**	58.5	59.7	60.9
4	Median	53.9	50.8	58.4	58.7	52.1	52.5	52.7
	Mean energy	55.4	55.9	58.7	**62.2**	54.7	54.6	55.3
5	Median	56.7	54.2	59.7	56.3	53.9	56.2	53.9
	Mean energy	58.8	57.3	**59.7**	58.9	55.6	55.9	54.4
Female Speaker								
6	Median	51.6	53.7	59.7	57.7	47.4	51.5	50.1
	Mean energy	55.7	55.7	**60.9**	59.4	50.8	54.4	52.8
7	Median	59.2	59.1	62.2	60.9	56.6	56.3	57.2
	Mean energy	59.3	59.4	62.3	**62.5**	57.3	56.3	57.6
8	Median	60.2	59.9	62.4	61.1	55.9	55.8	57.5
	Mean energy	61.8	61.4	63.3	**63.9**	56.3	57.1	58.1
9	Median	59.2	58.7	61.1	61.5	57.2	58.1	58.1
	Mean energy	59.2	59.8	62.5	**62.9**	57.3	57.9	59.2
10	Median	58.0	58.8	60.7	61.6	57.0	58.7	57.1
	Mean energy	58.2	58.8	61.8	**62.9**	57.3	57.9	57.2

* Bold data indicates the highest intensity value
* Underline data indicates the lowest intensity value

reciting the Quran are able to be solved. It is included the phoneme of *Sad* (ص) and *Seen* (س), which considered as frequent errors (see Table 1) that have repetitively been made by the majority of Malay speakers.

The median and mean energy of intensity reveal a significant difference in the phoneme of *Zain* (ز), which has higher energy values (**bold** value) as compared to the phoneme of *Sad* (ص) and phoneme of *Seen* (س) (less energy) (underline value), in both male and female respondents. The finding denotes a more abrupt closure of the consonant and vocal folds of stressed syllables, as supported by previous study [70]. The intensity value of phoneme *Seen* (س) (underline) is the lowest as compared to the other two phonemes of *Sad* (ص) and *Zain* (ز). This result supports the findings reported in [7], indicating the intensity value of phoneme *Seen* (س) should be the lowest. Meanwhile, the phoneme of *Zain* (ز) (bold) for Quranic accents of *Khalad* and *Khallaf-facet 1* have been considered as the stressed phoneme with a higher value of intensity, than that of phoneme *Sad* (ص). The higher energy of phoneme *Zain* (ز) also influences the stress differently due to the different regional of Quranic accents.

The experimental results based on intensity (energy and lexical stress) in this prosodic feature analysis, verified the dissimilarities between each CA phoneme used in Quranic accents recitation. Here, the apparent phonological diversity and uniqueness in the characteristics of CA phonemes, where the stressed syllables and accents for the particular consonants can be differentiated. According to this, the prosodic features are highly necessary to differentiate the Quranic accents, and identify misarticulated phoneme among Malay speakers during Quranic recitation. Hence, the results of Quranic accents obtained later can be considered as reliable. The accent information carried by prosodic are crucial for further classification process, due to enhance the modeling and recognition. From this analysis, the accentual influences, and prosodic impacts of the Malay speakers' recitation of Quranic accents have supported the study and literature review of the past research [12], [13], [14], on the issues of misarticulated phonemes of CA language (see Table 1). As a result, there are noticeable differences in prosody between each CA phoneme, as proven in acoustic analysis executed (see Table 6).

On the other hand, the result of pitch and duration have been presented in Table 7. The value of pitch and intensity are related and coincide between each other. Here, the phoneme *Seen* (س) achieved the highest pitch value, as compared to other phonemes of *Zain* (ز) and *Sad* (ص), but gained the lowest energy. Contradict with the result of phoneme *Zain* (ز), where it represented the highest energy value, but the pitch value was the lowest. Meanwhile, the duration for each of syllables (from three syllables), which represented the various Quranic accents has differed within the small margin of period only.

Noted from Table 7, the margin of period, which differentiate one accent to another is 0.092 seconds (highest), while the lowest is 0.004 seconds. Theoretically in Quranic recitation, these three syllables are combination of the consonants with short vowel of *fatHa*, that just needs to be pronounced by a duration of one motion/count only. In this case, none of the duration values listed from the table is statistically significant. Hence, the duration feature is unnecessary to be presented in this paper, because the result of duration was unable to show a significant difference between one accent to another (distinguish the Quranic accents). Due to this matter, the duration feature has not been highlighted and observed for the next assessments, although the duration (as one of element of prosodic features) is also involved with the computational process in this research work.

The comparison of gender based on three syllables is presented in Table 8, where the average values for each prosodic features are computed from the results as listed in Table 6 and Table 7. There is a clear trend of decreasing of the

TABLE 7. Pitch and duration values.

		Pitch (Hz) and Duration/Time values (sec)						
	Word	/sˤi ra: tˤa/ (صِرَاطَ)		/zi ra: tˤa/ (زِرَاطَ)		/si ra: tˤa/ (سِرَاطَ)		
	Syll.	/sˤi/ (ص)		/zi/ (ز)		/si/ (س)		
	Quran accent	Hafs	Bazzi	Khalad	Khallaf-facet 1	Khallaf-facet 2	Qunbul	Ruwais
Native Sp.	Pitch	159	170	150	143	**211**	197	196
	Time	0.32	0.29	0.25	0.26	0.25	0.24	0.27
Male Speaker								
1	Pitch	186	204	157	157	**224**	212	205
	Time	0.19	0.26	0.23	0.24	0.22	0.19	0.23
2	Pitch	121	123	111	114	126	124	**126**
	Time	0.27	0.25	0.27	0.28	0.35	0.28	0.31
3	Pitch	159	145	143	125	**167**	165	167
	Time	0.20	0.21	0.23	0.23	0.23	0.26	0.27
4	Pitch	169	182	161	136	**193**	184	184
	Time	0.20	0.21	0.17	0.25	0.23	0.22	0.23
5	Pitch	147	151	133	144	157	**159**	153
	Time	0.22	0.22	0.30	0.39	0.30	0.27	0.29
Female Speaker								
6	Pitch	235	242	213	219	**278**	250	244
	Time	0.17	0.16	0.21	0.22	0.20	0.20	0.19
7	Pitch	264	264	208	213	**272**	268	265
	Time	0.27	0.27	0.23	0.29	0.28	0.28	0.29
8	Pitch	258	257	252	240	275	282	**292**
	Time	0.21	0.23	0.26	0.21	0.23	0.24	0.23
9	Pitch	212	214	208	194	232	**238**	233
	Time	0.20	0.20	0.14	0.15	0.23	0.2	0.20
10	Pitch	215	213	204	193	238	**253**	238
	Time	0.20	0.19	0.14	0.15	0.23	0.24	0.21

* Bold data indicates the highest pitch value
* Underline data indicates the lowest pitch value

TABLE 8. Comparison of average values of energy, pitch, and duration.

		Energy (dB), Pitch (Hz), and Duration (sec)						
	Word	/sˤi ra: tˤa/ (صِرَاطَ)		/zi ra: tˤa/ (زِرَاطَ)		/si ra: tˤa/ (سِرَاطَ)		
	Syllable	/sˤi/ (ص)		/zi/ (ز)		/si/ (س)		
	Quran accents	Hafs	Bazzi	Khalad	Khallaf-facet 1	Khallaf-facet 2	Qunbul	Ruwais
Male	Median	53.7	55.1	59.2	58.6	51.9	51.4	51.5
	Mean energy	58.4	58.4	60.8	62.1	55.1	55.9	56.7
	Pitch	156	161	141	135	173	169	167
	Time	0.22	0.24	0.24	0.28	0.27	0.25	0.27
Female	Median	57.7	58.1	61.2	60.6	54.8	56.1	55.9
	Mean energy	58.9	58.9	62.2	62.4	55.8	56.7	56.9
	Pitch	237	238	217	212	259	258	255
	Time	0.22	0.22	0.20	0.20	0.24	0.23	0.23

* Blue data indicates the highest energy; and underline as the highest pitch
* Red data indicates the lowest energy; and underline as the lowest pitch

energy values for phoneme *Seen* (س), when the pitch value is high (values in blue and underline). Meanwhile, the energy values for phoneme *Zain*(ز) have gradually increased and reach to the maximum level when the pitch values reach to the lowest point. This situation had occurred for male and female speakers, but in a comparison between both genders, the values of pitch for female speakers is higher compared to male speakers. The expectation of the result is based on the research findings by [77], where the F0 value of a woman adult is about twice as high, while her vocal tract is about 15% shorter than a man.

b: ANOVA AND T-TEST ANALYSIS

The one-way ANOVA with Tukey's post hoc was performed to compare the energy, pitch, and duration between the Quranic accents, as shown in Table 9. However, only the energy value is presented in this paper, as an interpretation from the result of energy in Table 6.

TABLE 9. Energy (intensity) value.

Mean Energy						
Quranic accents		Mean Diff. (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower	Upper
Hafs	Bazzi	-0.071	1.136	1.00	-3.532	3.390
	Khalad	-2.823	1.136	0.18	-6.285	0.638
	Khallaf-1	-3.579*	1.136	0.03	-7.040	-0.117
	Khallaf-2	3.174	1.136	0.09	-0.286	6.636
	Qunbul	2.272	1.136	0.42	-1.189	5.733
	Ruwais	1.785	1.136	0.70	-1.676	5.246
Bazzi	Hafs	0.071	1.136	1.00	-3.390	3.532
	Khalad	-2.752	1.136	0.20	-6.213	0.709
	Khallaf-1	-3.508*	1.136	0.04	-6.969	-0.046
	Khallaf-2	3.245	1.136	0.08	-0.215	6.707
	Qunbul	2.343	1.136	0.38	-1.118	5.805
	Ruwais	1.856	1.136	0.66	-1.605	5.318
Kha-lad	Hafs	2.823	1.136	0.18	-0.638	6.285
	Bazzi	2.752	1.136	0.20	-0.709	6.213
	Khallaf-1	-0.755	1.136	0.99	-4.217	2.705
	Khallaf-2	5.998*	1.136	0.00	2.536	9.459
	Qunbul	5.095*	1.136	0.00	1.633	8.557
	Ruwais	4.608*	1.136	0.00	1.146	8.070
Khal -laf-1	Hafs	3.579*	1.136	0.03	0.117	7.040
	Bazzi	3.508*	1.136	0.04	0.046	6.969
	Khalad	0.755	1.136	0.99	-2.705	4.217
	Khallaf-2	6.753*	1.136	0.00	3.292	10.215
	Qunbul	5.851*	1.136	0.00	2.389	9.313
	Ruwais	5.364*	1.136	0.00	1.902	8.826
Khal -laf-2	Hafs	-3.174	1.136	0.09	-6.636	0.286
	Bazzi	-3.245	1.136	0.08	-6.707	0.215
	Khalad	-5.998*	1.136	0.00	-9.459	-2.536
	Khallaf-1	-6.753*	1.136	0.00	-10.21	-3.292
	Qunbul	-0.902	1.136	0.98	-4.364	2.559
	Ruwais	-1.389	1.136	0.88	-4.851	2.072
Qun-bul	Hafs	-2.272	1.136	0.42	-5.733	1.189
	Bazzi	-2.343	1.136	0.38	-5.805	1.118
	Khalad	-5.095*	1.136	0.00	-8.557	-1.633
	Khallaf-1	-5.851*	1.136	0.00	-9.313	-2.389
	Khallaf-2	0.902	1.136	0.98	-2.559	4.364
	Ruwais	-0.487	1.136	0.99	-3.948	2.974
Ru-wais	Hafs	-1.785	1.136	0.70	-5.246	1.676
	Bazzi	-1.856	1.136	0.66	-5.318	1.605
	Khalad	-4.608*	1.136	0.00	-8.070	-1.146
	Khallaf-1	-5.364*	1.136	0.00	-8.826	-1.902
	Khallaf-2	1.389	1.136	0.88	-2.072	4.851
	Qunbul	0.487	1.136	0.99	-2.974	3.948

*. The mean difference is significant at the 0.05 level

TABLE 10. T-Test values of energy, pitch, and duration (between male & female speakers).

	Levene's test for Equality of Variances			T-Test for Equality of Means						
				t	df	Sig. (2-tailed)	Mean Diff.	Std. Error Diff.	95% Confidence Interval of Difference	
		F	Sig.						Lower	Upper
Mean energy	Equal variance assumed	0.211	0.648	-0.754	68	0.454	-0.6148	0.8156	-2.2424	1.0127
	Equal variance not assumed			-0.754	65.546	0.454	-0.6148	0.8156	-2.2435	1.0138
Pitch	Equal variance assumed	0.030	0.863	-12.286	68	0.000	-81.7328	6.6524	-95.0076	-68.4580
	Equal variance not assumed			-12.286	67.502	0.000	-81.7328	6.6524	-95.0094	-68.4563
Dura-tion	Equal variance assumed	0.291	0.591	3.095	68	0.003	0.0325	0.0105	0.0115	0.0536
	Equal variance not assumed			3.095	67.350	0.003	0.0325	0.0105	0.0115	0.0536

The articulation of phonemes following *Khallaf-facet 1* accent resulted in significantly higher energy compared to other accents (except for *Khalad*). These Quranic accents represented the phoneme *Zain* (ز) from the word /zi ra: tˤa/, which achieved higher energy values, as compared to phoneme *Sad* (ص) and *Seen* (س) (lowest energy). Means, the result of this analysis has supported the evaluation and justification made from the previous acoustic analysis, as well as findings from [7]. In other hand, no significant difference is observed from the ANOVA analysis measured for pitch and duration between the accents. Thus, the ANOVA analysis for pitch and duration features is needless to be presented in this paper.

Another one-way ANOVA analysis was performed to compare the energy, pitch, and duration between the speakers of the same gender. But, the comparison of these features from the similar gender is considered irrelevant in distinguishing the Quranic accents. Therefore, the result obtained from this analysis is not highlighted here. Only the t-test value is presented in this paper (see Table 10), for a comparison between two different genders of speakers. Here, an independent sample t-test is performed to investigate the relation of energy, pitch, and duration between male and female speakers. Overall, no significant difference is observed between the male and female speakers in terms of energy, pitch, and duration. However, the female speakers are generally having a higher pitch and energy, with shorter duration compared to male speakers.

2) SPECTRAL AND PROSODIC FEATURES ANALYSIS

In this section, we examined the performance of various features for recognition of Quranic accents. The features are the combination of prosodic features (pitch, energy, duration, spectral-tilt) and spectral features[2] (MFCC-based measurement across the consonant-vowel or syllable nuclei).

As presented in Table 11, the performance results for feature extraction process are presented stage by stage, to provide a clear view of the significant improvement achieved by each feature and its combination. These different values obtained are significantly varied for each accent, which shows the mismatch between consonants and vowels. We can use this as an essential cue to the classifier for decision-making, especially for accent identification and classification. Noted that, the duration that represent as one of the prosodic features also involved with the computational process at all stages of assessments (except MFCC only). However, only pitch, energy and spectral-tilt were observed and highlighted in this paper for evaluation, and not for duration (based on previous justification). Whereas, the spectral features of MFCC is configured using 13-dimensional MFCC, extracted with energy, plus their delta and double delta are appended. The 39-dimensional vectors are subjected to cepstral mean and variance normalization.

TABLE 11. The percent accuracy of recognition based on prosodic and spectral features using MFCC & GMM.

Feature extraction	Classification (Testing-phase)				
	Classi-fier	Train-set (%)		Test-set (%)	
		Acc	EER	Acc	EER
MFCC	GMM	84.22	15.8	74.4	25.6
MFCC + Pitch		87.72	12.3	77.9	22.1
MFCC + Pitch + Energy		89.061	10.9	80.1	19.9
MFCC + Pitch + Energy + Spectral-tilt		89.697	10.3	81.7	18.3

* Underline data indicates the best results

In this experiment, we implemented the conventional classification technique of GMM with 64 mixture components for accent modeling, due to estimate the weight of the mean vector for each Quranic accent, as well as the mixture of the weights from the training of Quranic pronunciation. The results obtained from Table 11 show a significant improvement of 7.303% (test-set) and 5.477% (train-set). It was rendered by the use of spectral features (MFCC) integrated with prosodic features (pitch, energy, duration, and spectral-tilt), which are capable to produce better recognition results, as compared to the previous ASR research using spectral features of MFCC only [21], [22], [23], [29], [31]. Inspired by this result, therefore, we adopt this feature extraction technique for the rest of the experiment after this. We believed the combination of spectral and prosodic features, that extracted from the speech signal can provide sufficient information for supporting the further process of classification of Quranic accents recitation recognition.

[2]Pitch extracted from spectral features (Fig. 3) excluded for evaluation, since the prosodic features already extracted the long speech segments.

B. BACK-END PROCESSING (GMM-UBM)

As previously discussed in Section II-B, the crucial issue in the Quranic accents recitation involved the misarticulated phonemes, which are influenced by the Malay colloquial dialect. Based on Table 1, there are certain CA phonemes have been categorized as Quranic accents phonemes, which has a close articulation sounds for each other, according to Malay language pronunciation. But in fact, the articulation point has differed in CA language. The slightest differences of CA phonemes between each of accent in Quran has differentiated the Quranic accent's reading style and the way on how they supposed to be read (i.e., seven Quranic accents). For discriminating an accent under the presence of closely articulated sounds, we need a considerable amount of training data for developing an accent model. The modeling method ought to have a great deal of mixture components, in order to identify the slight variation presented from each Quranic accent. Collecting and dealing with a considerable amount of data for each class, to train an accent model with a large number of mixture components might not be possible if we implemented the conventional classification technique of GMM. In other word, the implementation of GMM in Quranic accents is not practically effective, because it is prone to overfitting.

In this work, we proposed to implement the GMM-UBM classification technique for developing the accent models. Using the modeling technique of GMM-UBM, the speech data of 4 speakers (see Section III-B 2(b)) from all classes of Quranic accents is pooled, in order to develop a universal background model with a large number of mixture components. Then, this UBM model is adapted to all classes of Quranic accents, using the training data from other 6 speakers. In the present approach, we used 10 EM iterations to train the universal background model for the GMM-UBM system with 512 mixture components. Here, the UBM model is developed using 116-140 minutes of speech data, comprising of all the Quranic accents. This UBM model is adapted to all classes to build Quranic accents models with 512. During the testing phase, the speech data from 4 speakers (testing dataset) for each Quranic accent, that consist of both male and female speakers are used to test the developed Quranic accents models. The performance of GMM-UBM is discussed in the following sub-section.

1) COMPARISON TO OTHER MODELS

In our initial experiment, we compared the performance of our proposed classification technique of GMM-UBM with other modeling techniques. Specifically, the other techniques are the k-NN [29], [78], GMM [21], [22], [23], [26], [30], and GMM-iVector [79]. The goal is to compare the performance of these different identification methods using the similar datasets (as listed in Table 3) and front-end processing, as discussed earlier. These different modeling techniques are interesting to compare, because they represent different ways of modeling, due to determine the best classification technique that capable of differentiating the Quranic accents and classify the Quranic accents effectively.

TABLE 12. Overall results of recognition between classification.

Feature extraction	Classification (Testing-phase)				
	Classifier	Train-set (%)		Test-set (%)	
		Acc	EER	Acc	EER
MFCC + Pitch + Energy + Spectral-Tilt	k-NN	88.397	11.603	79.470	20.53
	GMM	89.697	10.303	81.713	18.287
	GMM – iVector	89.898	10.102	82.166	17.834
	GMM - UBM	90.255	9.745	86.148	13.852

* Underline data indicates the best results

As for GMM-iVector, the training data that comprises all seven Quranic accents are utilized in training the hyper-parameter of the i-vector system, UBM, and T-matrix. The UBM model of 512 Gaussian components is trained using the EM algorithm. The total variability subspace of dimension 400 is applied for the i-vector. In the i-vector approach, the GMM supervector of each Quranic accent utterance is computed, where MLE of the total variability subspace (T-matrix) is calculated. Here, T-matrix is learned from the EM, and used to estimate i-vector from its posterior distribution on the Baum–Welch statistics, which extracted from the utterance using the UBM. Unlike the GMM-UBM (uses feature vectors), i-vectors are used to represent the model and test segments. The dimensionality of the i-vectors is reduced through Linear Discriminant Analysis (LDA) using the Fisher criterion. The aim is to compute a linear transformation that maximized the between-accent variations, while minimizing the intra-accent variations. Then, gaussian Probabilistic LDA (PLDA) modeling is performed, after whitening, mean and length (normalized) are executed. Lastly, the identification result from the system is given by calculating the log-likelihood ratio (LLR) and verification scores, where the system performance is then presented by accuracy.

The general trend shown by Table 12, Table 13-(a) and Table 13-(b) is, as the values of EER decreased, a significant increment of performance for accuracy occurs in all classifications. In this experiment, we made a comparative analysis to compare the efficiency indices of four different models, as shown in Table 12. To determine the best results and classifier among them, we use a 10-fold cross-validation technique to evaluate the trained classifier models. Here, the GMM-UBM demonstrate the highest efficiency, with average accuracy is 86.15% (test-set) and 90.26% (train-set), outperforming the other models by 1-7%. Besides, the GMM-UBM performance is recorded by 6.678% (test-set) and 1.858% (train-set) improvement, as compared to conventional k-NN. Also, the average accuracy indices calculated for GMM-UBM are higher than that of conventional GMM, with an increment of 4.435% (test-set) and 0.558% (train-set). The results show that the GMM-UBM models have produced 13.85% of EER (test-set) and 9.745% of EER (train-set). The GMM-UBM has also performed very well, as compared to

TABLE 13. (a) Average efficiency indices of classification technique (test-set). (b) Average efficiency indices of classification technique (train-set).

(a)

Efficiency Indices	k-NN	GMM	GMM-iVector	GMM-UBM
Accuracy (%)	79.470	81.713	82.166	86.148
Equal Error Rate (EER)(%)	20.53	18.287	17.834	13.852

* Underline data indicates the best results

(b)

Efficiency Indices	k-NN	GMM	GMM-iVector	GMM-UBM
Accuracy (%)	88.397	89.697	89.898	90.255
Equal Error Rate (EER)(%)	11.603	10.303	10.102	9.745

* Underline data indicates the best results

GMM-iVector, as the former improved the average accuracy indices by 3.982% (test-set) and 0.357% (train-set). Thus, according to efficiency indices obtained based on percent of accuracy and error rate (EER), it can be said that the GMM-UBM model has outperformed other classification tested in this research work.

The overall result shows some relevant observations. All evaluation merits recorded, demonstrate the superiority of the GMM-UBM model using the same data and same feature extraction technique (combined prosodic and spectral features), in distinguishing the different Quranic accents, as compared to that of k-NN, GMM-iVector and conventional GMM. We summarize the results on performance of all models in Table 13-(a) and Table 13-(b). From the practical perspective and consideration of the results obtained, it can be said that the GMM-UBM is an appropriate model to be used in the machine learning, particularly to distinguish the different Quranic accents. The results show that, a system using the model is able to recognize the different Quranic accents from the various features extracted from the test speech database. This information is important to help draw a comparative conclusion based on the performance.

The implementation of GMM-UBM through adaptation method able to increase the level of effectiveness in ASR modeling for Quranic accents. Adaptation also allowed the Malay speakers to update their correct articulation of CA phonemes in Quran, to prevent from confusion and mis-articulated phonemes (blue and red color of phonemes as highlighted in Table 4) later on, during testing stage. Hence, any slight differences from CA phonemes can be differentiated and identified clearly, based on respective Quranic accents.

As compared to conventional GMM classification, an accent model is computed independently by using the EM algorithm. Commonly, the multi-variate of GMM could suffer from the overfitting problem. This issue is occurred when the model complexity is high, under the presence of singularity. Moreover, the implementation of this classification becomes complicated when the size of training data is larger, and thus processing time become very time consuming, especially with a large number of mixture components in GMM. The GMM issue become complicated, where the GMM algorithm having lack of systematic way in managing a large training data with multiple accents of Quran to be modeled with different classes of Quranic accents. Without the adaptation function like GMM-UBM, the model created through GMM becomes less sensitive/robust, where it is difficult to differentiate between each of CA phoneme and prone to error (testing). We believed that the absence of this function is considered as a major drawback for GMM algorithm.

2) CONFUSION MATRIX (GMM-UBM)

The confusion matrix developed for the GMM-UBM model in the proposed system presents the best value that signifies the correct Quranic accents in the testing phase. Here, the matrix depicts the confusion faced by a particular classifier in selecting the correct pattern based on detection of similarities with the training set of Quranic accents in the model. The confusion matrices developed for Quranic accents are tabulated in Table 14-(a) and Table 14-(b).

In this case, the results of validation accuracy for GMM-UBM classification, which is presented based on a 10-folds confusion matrix, are depicted in both tables. Here, the values underline (in a diagonal position) represent the accuracy of the correct Quranic accents. By using the proposed GMM-UBM classifier, the system achieved classification accuracies up to 98.4% for train-set, while the test-set achieved up to 96.8% classification accuracy. It is observed that, the proposed system from the train-set (see Table 14-(b)) achieved better results than the test-set (see Table 14-(a)), for all the Quranic accents classes. The system achieved a significant improvement, especially in identifying the Quranic accent of *Qunbul* (98.4%) and *Hafs* (90.7%) from the train-set, which is presented in Table 14-(b).

3) EFFICIENCY INDICES (GMM-UBM)

Meanwhile, Table 15-(a) and Table 15-(b) show the results obtained after using the 10-fold cross validation technique. The data from both tables present the efficiency indices for identification of Quranic accents using the GMM-UBM classification in both test-set and train-set, respectively. We presented the experimental results based on four common performance measures, known as of precision (p), sensitivity, specificity, and F-Score ($F1$), for each set of Quranic accents. From the data presented in both tables, we can see that the performance of the GMM-UBM model for the majority of Quranic accents has achieved the optimal result. It is apparent from these tables that most of Quranic accents have accomplished above 80% of efficiency, which considered as ideal results.

Finally, the experimental results as presented in Table 13-(a) and Table 13-(b), Table 14-(a) and Table 14-(b), as well as Table 15-(a) and Table 15-(b) are quite revealing in several ways. First, unlike the other tables, the confusion matrix and efficiency indices for each of Quranic accent

TABLE 14. (a) The confusion matrix based on prosodic & spectral features and GMM-UBM (test-set) (%). (b) The confusion matrix based on prosodic & spectral features and GMM-UBM (train-set) (%).

(a)

Quranic accents	(MFCC + Pitch + Energy + Spectral-Tilt) and GMM-UBM						
	Hafs	Khalad	Khallaf_1	Khallaf_2	Bazzi	Qunbul	Ruwais
Hafs	78.7	4.68	5.36	0	7.81	0	3.39
Khalad	0	84.3	7.69	0	2.31	0	5.71
Khallaf_1	0	0	96.8	3.21	0	0	0
Khallaf_2	0	0	0	86.9	0.09	7.81	5.13
Bazzi	6.82	0	0	0	88.4	4.79	0
Qunbul	3.66	4.27	0	4.39	0	87.7	0
Ruwais	0	0	1.25	0	9.29	9.29	80.2

* Underline data indicates the best results/correctly classified

(b)

Quranic accents	(MFCC + Pitch + Energy + Spectral-Tilt) and GMM-UBM						
	Hafs	Khalad	Khallaf_1	Khallaf_2	Bazzi	Qunbul	Ruwais
Hafs	90.7	0	5.07	0	0	4.21	0
Khalad	8.65	84.8	0	2.32	0	4.20	0
Khallaf_1	0	0	92.3	0	0	0	7.68
Khallaf_2	6.69	0	0	89.1	0	4.20	0
Bazzi	8.65	0	0	0	86.6	4.74	0
Qunbul	0	0	0	0	1.61	98.4	0
Ruwais	2.55	0	3.45	0	0	4.19	89.8

* Underline data indicates the best results/correctly classified

using GMM-UBM classification are presented. Second, the different types of classification models are tested, including the different GMM-based classification models, such as conventional GMM, k-NN and GMM-iVector. Although GMM has been the dominant approach used in ASR research by Arabic and CA researchers, the model still suffers with overfitting problem especially when the model is highly complex. In such cases, the GMM-UBM performs better than other classification techniques, including conventional GMM, as observed in the efficiency indices obtained for both test-set and train-set (see Table 13-(a) and Table 13-(b)). In GMM-UBM approach, we initially develop a large model by using data from all seven Quranic accents to be adapted to all accents, in order to develop the accent-based Quranic recitation models. Therefore, the combined features of spectral (MFCC) and prosodic in the GMM-UBM classification have produced a model, which considered ideal and reliable for ASR recognition with accent identification, to assist and recognize the Quranic accents verse recitation.

TABLE 15. (a) Efficiency indices for Quranic accents (test-set). (b) Efficiency indices for Quranic accents (train-set).

(a)

Efficiency Indices	Quranic accents (Qira'at)						
	Hafs	Khalad	Khallaf_1	Khallaf_2	Bazzi	Qunbul	Ruwais
Precision (%)	88.3	90.4	87.1	91.9	81.9	80.0	84.9
Sensitivity (%)	78.8	84.3	96.8	86.9	88.4	87.7	80.2
Specificity (%)	98.3	98.5	96.6	98.5	96.8	96.4	97.6
F-Score (%)	83.2	87.2	91.7	89.4	85.0	83.7	82.5

(b)

Efficiency Indices	Quranic accents (Qira'at)						
	Hafs	Khalad	Khallaf_1	Khallaf_2	Bazzi	Qunbul	Ruwais
Precision (%)	79.2	96.9	91.6	97.5	98.2	82.0	92.1
Sensitivity (%)	90.7	84.8	92.3	89.1	86.6	98.4	89.8
Specificity (%)	96.0	99.5	98.6	99.6	99.7	96.4	98.7
F-Score (%)	84.6	90.4	91.9	93.1	92.0	89.5	90.9

V. CONCLUSION AND FUTURE WORK

In this study, we have introduced a dataset for Quranic accents and an ASR methodology for CA language. A summary of the conclusion are as follows:

- A new database of Quranic accents has been developed and evaluated through a series of experiments. However, the quality of the database could not be evaluated via comparison, as there is no standard database of Quranic accents available yet.
- The feature extraction in the proposed methodology incorporated prosodic (instead of commonly used spectral features only) and spectral features. The prosodic features carry traits of accents, which contribute to the robustness of the ASR system for the CA language.
- Using the adaptation method from UBM, is greatly simplifies the training and brings it into a much more effective way of training the models, which then allows for a fast-scoring technique during testing. As compared to GMM, the modeling and recognition process can be quite time-consuming for larger databases and could cause overfitting problems.
- We compared the performance results of the well-known GMM with other classification techniques, such as k-NN, GMM-iVector, and GMM-UBM. After executing the proposed GMM-UBM classification, the results have significantly outperformed other classification techniques in all cases. We also observed the improvements when the prosodic been used together

with spectral features (extract speech and accent info) from the beginning till the end, versus the common approach of spectral features and conventional classification for training and testing.

- Finally, we showed our Quranic accents recognizer outperforms other classification techniques, in similar conditions. Here, the result of accuracy for GMM-UBM is the best with 86.15% (test-set) and 90.26% (train-set), whereby EER is 13.8% and 9.7%, respectively. The results obtained after the implementation of GMM-UBM had met our expectations, which behaved fairly and precisely for both female and male recitations of Malay speakers. Thus, we believed the level of effectiveness of the developed system in detecting the differences of CA phonemes based on the Quranic accents has improved. Moreover, the use of prosodic features enables the classification process to identify the differences between CA phonemes, with the misarticulated phonemes problem faced by Malay speakers.

As demand for self-learning tool of Quranic accents among non-Arabic speakers increases, the combination of prosodic and spectral features for classification by GMM-UBM embarks on a new method in identifying the Quranic accents. This study should be optimized and improved further in the future, with a recognition technique specifically for the Quranic accents recitation. Therefore, in the future, we will extend this work for other non-native Arabic speakers and cover other chapters of the Quran. Means, we will expand this research work by applying with a better-quality audio of speech samples and larger size of speech database, where the amount of training data will be increased for each Quranic accent. Other than enlarging the chapters of the Quran, expanding to other types of Quranic accents, through replication procedure is also highly beneficial indeed. All the trials, evaluations, and analysis conducted in this research have led us to the different ideas for future work and improvement, especially in developing a database for various Quranic accents.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their valuable comments, and the research team from the Centre of Quranic Research (CQR), University of Malaya, for their hospitality, supports, and collaboration.

REFERENCES

[1] N. F. Chen, S. W. Tam, W. Shen, and J. P. Campbell, "Characterizing phonetic transformations and acoustic differences across English dialects," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 22, no. 1, pp. 110–124, Jan. 2014.

[2] D. T. Grozdic and S. T. Jovicic, "Whispered speech recognition using deep denoising autoencoder and inverse filtering," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 12, pp. 2313–2322, Dec. 2017.

[3] M. Liu, B. Xu, T. Hunng, Y. Deng, and C. Li, "Mandarin accent adaptation based on context-independent/context-dependent pronunciation modeling," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Istanbul, Turkey, Jun. 2000, pp. 1025–1028.

[4] M. N. Abdullah, *A Clear Path of 10 Quranic Accents of Recitation (Manhaj Qiraat 10)*. Kuala Lumpur, Malaysia: Pustaka Salam Sdn. Bhd., 2013, pp. 1–13.

[5] M. F. M. Azali, M. A. Samsungei, and M. H. Othman, "A brief note of the ten Qiraat according to Al-Syatibiyah and Al-Durrah (Nota ringkas Qiraat sepuluh menurut Al-Syatibiyah dan Al-Durrah)," *Al-Quran & Qira'at Studies*. Malaysia: Darul Quran, JAKIM, 2016, pp. 2–19.

[6] M. R. M. A'Tarahim, Y. N. Hafizi, M. Y. Zulkifli, D. Normadiah, M. I. M. F. Hakimi, Y. Sofyuddin, A. W. A. Hisham, and S. A. Zahid, "Riwayah of Hafs and Warsh recitation methods: The case of Maqam Ibrahim," *Pertanika J. Soc. Sci. Hum.*, vol. 25, pp. 103–108, May 2017.

[7] Y. Alotaibi and A. Meftah, "Review of distinctive phonetic features and the Arabic share in related modern research," *Turkish J. Electr. Eng. Comput. Sci.*, vol. 21, no. 5, pp. 1426–1439, 2013.

[8] K. Abu-bakar and M. F. Abdullah, "Word stress of Arabic as a foreign language amongst Malay speakers (Tekanan perkataan Arab sebagai bahasa asing dlm kalangan penutur Melayu)," *GEMA Online J. Lang. Stud.*, vol. 18, no. 1, pp. 87–105, Feb. 2018.

[9] N. A. Ramli, C. R. Mezah, and Y. N. Thai, "Malay students' mastery of voice stress in saying the Arabic words from the aspect of intensity (Penguasaan Pelajar Melayu Terhadap Tekanan Suara Menyebut Perkataan Arab dari Sudut Intensiti)," *J. Humanit.*, vol. 14, no. 1, pp. 1–14, Nov. 2016.

[10] Z. M. Don, G. Knowles, and J. Yong, "How words can be misleading: A study of syllable timing and 'stress' in Malay," *J. Linguistic Anthropol.*, vol. 3, no. 2, pp. 66–81, Aug. 2008.

[11] S. S. Hassan and M. A. Zailaini, "The forms of Quranic recitation errors by students in an IPTA (Bentuk-bentuk kesalahan bacaan al-Quran pelajar di sebuah IPTA)," *O-JIE, Online J. Islamic Educ.*, vol. 3, no. 2, pp. 1–9, Jul. 2015.

[12] M. N. M. Hasbullah, "Analysis of Quranic pronunciation errors among secondary school students (Analisis kesilapan sebutan bahasa Al-Quran di kalangan pelajar sekolah menengah)," Ph.D. dissertation, Dept. Al-Quran Al-Hadith, Malaya Univ., Kuala Lumpur, Malaysia, 2001.

[13] N.-A. Abdul-Kadir and R. Sudirman, "Difficulties of standard Arabic phonemes spoken by non-Arab primary school children based on formant frequencies," *J. Comput. Sci.*, vol. 7, no. 7, pp. 1003–1010, Jul. 2011.

[14] A. M. K. Mannan, "Optimization of Arabic speech recognition for non-native speakers using diverse training corpora," M.S. thesis, Dept. Comput. Syst. Tech., Malaya Univ., Kuala Lumpur, Malaysia, 2013.

[15] N. M. Rahimi, H. Baharudin, Y. Ghazali, K. S. M. The, and M. A. Embi, "Learning the Arabic consonants through Malay accent (Pembelajaran konsonan Arab mengikut pelat bahasa Melayu)," *GEMA Online J. Lang. Stud.*, vol. 10, no. 3, pp. 1–14, Sep. 2010.

[16] H. Dahan, and A. Mannan, "Arabic speech pronunciation recognition and correction using Automatic Speech Recognizer (ASR)," in *Proc. 6th Int. Technol., Educ. Dev. Conf. (INTED)*, Valencia, Spain, 2012, pp. 4009–4016.

[17] I. Abdullah, personal communication/interview, Jan. 2017.

[18] M. L. Ibrahim, personal communication/interview, Jun. 2015.

[19] M. A. Shahin, B. Ahmed, and K. J. Ballard, "Automatic classification of unequal lexical stress patterns using machine learning algorithms," in *Proc. IEEE Spoken Lang. Technol. Workshop (SLT)*, Miami, FL, USA, Dec. 2012, pp. 388–391.

[20] N. J. Ibrahim, M. Y. I. Idris, Z. Razak, and N. N. A. Rahman, "Automated Tajweed checking rules engine for Quranic learning," *Multicultural Educ. Technol. J.*, vol. 7, no. 4, pp. 275–287, Nov. 2013.

[21] T. S. Gunawan, N. A. M. Saleh, and M. Kartiwi, "Development of Quranic reciter identification system using MFCC and GMM classifier," *Int. J. Elect. Comput. Eng. (IJECE)*, vol. 8, no. 1, pp. 372–378, Feb. 2018.

[22] M. O. M. Khelifa, M. Belkasmi, O. Yahya, and A. Yousfi, "Strategies for implementing an optimal ASR system for Quranic recitation recognition," *Int. J. Comput. Appl.*, vol. 172, no. 9, pp. 35–41, Aug. 2017.

[23] M. M. A. Baig, S. A. Qazi, and M. B. Kadri, "Discriminative training for phonetic recognition of the holy Quran," *Arabian J. Sci. Eng.*, vol. 40, no. 9, pp. 2629–2640, Jun. 2015.

[24] M. S. Abdo and A. H. Kandi, "Semi-automatic segmentation system for syllables extraction from continuous Arabic audio signal," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 1, pp. 535–540, 2016.

[25] T. AlTalmas, W. Sediono, N. N. W. N. Hashim, S. Ahmad, and S. Khairuddin, "Analysis of two adjacent articulation Quranic letters based on MFCC and DTW," in *Proc. 7th Int. Conf. Comput. Commun. Eng. (ICCCE)*, Kuala Lumpur, Malaysia, Sep. 2018, pp. 187–191.

[26] Y. Afrillia, H. Mawengkang, M. Ramli, and R. P. Fhonna, "Performance measurement of Mel frequency Ceptral Coefficient (MFCC) method in learning system of Al-Qur'an based in *Nagham* pattern recognition," in *Proc. Int. Conf. Inform. Commun. Technol. (IconICT)*, Medan, Indonesia, 2017, pp. 207–212.

[27] E. Mourtaga, A. Sharieh, and M. Abdallah, "Speaker independent Quranic recognizer based on Maximum Likelihood Linear Regression," *World Acad. Sci. Eng. Technol.*, vol. 20, no. 36, pp. 61–67, 2007.

[28] N. Kamarudin, S. A. R. Al-Haddad, S. J. Hashim, M. A. Nematollahi, and A. R. B. Hassan, "Feature extraction using spectral centroid and Mel Frequency Cepstral Coefficient for Quranic accent automatic identification," in *Proc. IEEE student Conf. Res. Develop.*, Dec. 2014, pp. 1–6.

[29] N. Kamarudin, S. A. R. Al-Haddad, M. A. Abushariah, S. J. Hashim, and A. R. B. Hassan, "Acoustic echo cancellation using adaptive filtering algorithms for Quranic accents (Qiraat) identification," *Int. J. Speech Technol.*, vol. 19, no. 2, pp. 393–405, Jun. 2016.

[30] N. Kamarudin, S. Al-Haddad, A. Khmag, A. B. Hassan, and S. J. Hashim, "Analysis on Mel Frequency Cepstral Coefficients and Linear Predictive Cepstral Coefficients as feature extraction on automatic accents identification," *Int. J. Appl. Eng. Res.*, vol. 11, no. 11, pp. 7301–7307, Jun. 2016.

[31] K. Nahar, H. Al-Muhtaseb, W. Al-Khatib, M. Elshafei, and M. Alghamdi, "Arabic phonemes transcription using data driven approach," *Int. Arab J. Inf. Technol.*, vol. 12, no. 3, pp. 237–245, May 2015.

[32] A. Mahmood, M. Alsulaiman, and G. Muhammad, "Automatic speaker recognition using multi-directional local features (MDLF)," *Arabian J. Sci. Eng.*, vol. 39, no. 5, pp. 3799–3811, May 2014.

[33] H. Tabbal, W. Al-Falou, and B. Monla, "Analysis and implementation of an automated delimiter of 'Quranic' verses in audio files using speech recognition techniques," in *Robust Speech Recognition and Understanding*. Vienna, Austria: IntechOpen, 2007, pp. 351–362.

[34] H. Behravan, "Dialect and accent recognition," M.S. thesis, School Comput., Fac. Sci., Forestry Technol., Univ. Eastern Finland, Kuopio, Finland, 2012.

[35] H. Li, "Deep learning for natural language processing: Advantages and challenges," *Nat. Sci. Rev.*, vol. 5, no. 1, pp. 24–26, Jan. 2018.

[36] N. B. Chittaragi, A. Prakash, and S. G. Koolagudi, "Dialect identification using spectral and prosodic features on single and ensemble classifiers," *Arabian J. Sci. Eng.*, vol. 43, no. 8, pp. 4289–4302, Nov. 2017.

[37] N. J. Ibrahim, M. Y. I. Idris, M. Y. Z. M. Yusoff, N. N. A. Rahman, and M. I. Dien, "Robust feature extraction based on spectral and prosodic features for classical Arabic accents recognition," *Malaysian J. Comput. Sci.*, Special no. 3, pp. 46–72, Dec. 2019.

[38] F. Barkatulla. (Oct. 2013). *The Importance of Tajweed*. Accessed: Nov. 2, 2014. [Online]. Available: https://www.al-islam.org/

[39] A. Mohammed and M. S. Sunar, "Verification of Quranic verses in audio files using speech recognition techniques," in *Proc. 1st Int. Conf. Recent Trends Inform. Commun. Technol. (IRICT)*, Johor, Malaysia, 2014, pp. 370–381.

[40] A. Z. S. A. Hadi, M. R. Ramlee, and N. M. Amin, "Enhancing teaching and learning methodology with computing visualization in studies of Qiraat (Malaysia)," *Int. J. Acad. Res. Bus. Soc. Sci.*, vol. 8, no. 2, pp. 823–835, Feb. 2018.

[41] M. O. AlQahtany, Y. A. Alotaibi, and S.-A. Selouani, "Analyzing the seventh vowel of classical Arabic," in *Proc. Int. Conf. Natural Lang. Process. Knowl. Eng. (NLP-KE)*, Dalian, China, Sep. 2009, pp. 1–7.

[42] M.-T. Luong, P. Nakov, and M.-Y. Kan, "A hybrid morpheme-word representation for machine translation of morphologically rich languages," 2019, *arXiv:1911.08117*.

[43] M. Maamouri, A. Bies, and S. Kulick, "Diacritization: A challenge to Arabic Treebank annotation and parsing," in *Proc. Int. Conf. Challenge Arabic NLP/MT*, London, U.K., 2006, pp. 35–47.

[44] S. M. Zaini. (May 30, 2019). *The Isyak Prayer at Puncak Alam Mosque Last Monday is Void (Solat Isyak di Masjid Puncak Alam isnin lalu tak sah)*. MyMetro. Accessed: May 31, 2019. [Online]. Available: http://www.hmetro.com.my/

[45] S. M. Zaini and N. C. Noh. (May 30, 2019). *The Imam Abandon 1 Verse While Prayer, and the Followers were Asked to Replace the Isyak Prayer (Imam Tertinggal Sepotong Ayat Ketika Solat, Makmum Diminta Qada Solat Isyak)*. Berita Harian. Accessed: May 31, 2019. [Online]. Available: http://www.bharian.com.my

[46] I. Zahid, "Prosody in teaching and learning of Malay language: The formation of personality and identity (Prosodi dalam P&P bahasa Melayu: Pembentukan sahsiah dan jati diri)," *J. Lang. Inst. Lang. Literature Brunei*, vol. 13, no. 23, pp. 3–17, 2011.

[47] B. A. Bakar, Z. M. Yusoff, and M. A. Norasid, "A clear path (Manhaj) of Al-Quran on Da'wah to adolescents: A bibliomatric study (Manhaj al-Quran mengenai Dakwah terhadap remaja: Satu kajian bibliomatrik)," *Al-'Ulwan J.*, vol. 3, no. 1, pp. 70–88, Jan. 2018.

[48] H. Hussin and M. Ismail, "The effectiveness of Al-Matien method in learning Al-Quran recitation (Keberkesanan kaedah Al-Matien dalam pembelajaran tilawah Al-Quran)," *Al-Turath, J. Al-Quran Al-Sunnah*, vol. 3, no. 2, pp. 1–8, Dec. 2018.

[49] H. Kasan, M. F. Mustafa, S. S. Haimi, and U. Faruk, "Interaction factors with Al-Quran in the appreciation process of students' religious lives in UKM (Faktor interaksi dengan Al-Quran dalam proses penghayatan kehidupan beragama pelajar-pelajar UKM)," *Jurnal Islam dan Masyarakat Kontemporari*, vol. 15, no. 1, pp. 84–93, Jul. 2017.

[50] S. M. Abdou, S. E. Hamid, M. Rashwan, A. Samir, O. Abdel-Hamid, M. Shahin, and W. Nazih, "Computer aided pronunciation learning system using speech recognition techniques," in *Proc. 9th Int. Conf. Spoken Lang. Process. INTERSPEECH (ICSLP)*, Pittsburgh, PA, USA, Sep. 2006, pp. 849–852.

[51] S. M. Abdou and M. Rashwan, "A computer aided pronunciation learning system for teaching the holy Quran recitation rules," in *Proc. IEEE/ACS 11th Int. Conf. Comput. Syst. Appl. (AICCSA)*, Doha, Qatar, Nov. 2014, pp. 543–550.

[52] M. S. Ridhwan, A. M. Zeki, and A. Olowolayemo, "Differential Qiraat processing applications using Spectrogram voice analysis," in *Proc. Int. Conf. Data Mining, Multimedia, Image Process. Appl. (ICDMMIPA)*, Kuala Lumpur, Malaysia, 2016, pp. 30–38.

[53] C. M. Bishop, *Pattern Recognition and Machine Learning*, vol. 128. Heidelberg, Germany: Springer, 2006, pp. 1–674.

[54] E. Wong and S. Sridharan, "Methods to improve Gaussian mixture model based language identification system," in *Proc. 7th Int. Conf. Spoken Lang. Process. (ICSLP)*, Denver, CO, USA, Sep. 2002, pp. 1–4.

[55] W. Shen and D. Reynolds, "Improved GMM-based language recognition using constrained MLLR transforms," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Las Vegas, NV, USA, Mar. 2008, pp. 4149–4152.

[56] T. Setiyorini and R. T. Asmono, "Implementation of gain ratio and K-Nearest Neighbor for classification of student performance," *Jurnal Pilar Nusa Mandiri*, vol. 16, no. 1, pp. 19–24, Mar. 2020.

[57] G. Ulumudin, A. Adiwijaya, and M. Mubarok, "A multilabel classification on topics of Qur'anic verses in English translation using K-Nearest Neighbor method with Weighted TF-IDF," *J. Phys., IOP Conf. Ser.*, vol. 1192, no. 1, pp. 1–7, 2019.

[58] J. Li, W. Dai, F. Metze, S. Qu, and S. Das, "A comparison of deep learning methods for environmental sound," 2017, *arXiv:1703.06902*.

[59] J. Gomes and M. El-Sharkawy, "i-Vector algorithm with Gaussian mixture model for efficient speech emotion recognition," in *Proc. Int. Conf. Comput. Sci. Comput. Intell. (CSCI)*, Las Vegas, NV, USA, Dec. 2015, pp. 476–480.

[60] L. Li, Y. Chen, D. Wang, and C. Zhao, "Weakly supervised PLDA training," 2016, *arXiv:1609.08441*.

[61] A. Poddar, M. Sahidullah, and G. Saha, "Performance comparison of speaker recognition systems in presence of duration variability," in *Proc. Annu. IEEE India Conf. (INDICON)*, New Delhi, India, Dec. 2015, pp. 1–6.

[62] A.-K. L. Al-Fatawi. (Mar. 11, 2019). *Law of Imam While Reciting Al-Fatihah or Other Verses in Different Quranic Accents Style (Hukum Imam Membaca Al-Fatihah Atau Ayat-Ayat Yang Lain Dengan Qiraat Yang Pelbagai)*. Kuala Lumpur, Malaysia: Mufti of Federal Territory of Malaysia. Accessed: May 31, 2019. [Online]. Available: https://muftiwp.gov.my/en/artikel/al-kafi-li-al-fatawi/

[63] C. Moler and J. Little, "A history of MATLAB," in *Proc. ACM Program. Lang.*, vol. 4, Jun. 2020, pp. 1–67.

[64] P. Ladefoged, *A Course in Phonetics*, 3rd ed. Fort Worth, TX, USA: Harcourt, 1993.

[65] S. Sinha, S. S. Agrawal, and A. Jain, "Dialectal influences on acoustic duration of Hindi phonemes," in *Proc. Int. Conf. Oriental (COCOSDA), Conf. Asian Spoken Lang. Res. Eval. (CASLRE)*, Nov. 2013, pp. 1–5.

[66] P. Boersma and D. Weenink. *Praat: Doing phonetics by Computer*. Institute of Phonetic Sciences, Amsterdam, Netherlands. Accessed: May 5, 2014. [Online]. Available: https://www.praat.org/

[67] M. Y. El Amrani, M. M. H. Rahman, M. R. Wahiddin, and A. Shah, "Building CMU sphinx language model for the holy Quran using simplified Arabic phonemes," *Egyptian Informat. J.*, vol. 17, no. 3, pp. 305–314, Nov. 2016.

[68] N. Singh, R. A. Khan, and R. Shree, "MFCC and prosodic feature extraction techniques: A comparative study," *Int. J. Comput. Appl.*, vol. 54, no. 1, pp. 9–13, Sep. 2012.

[69] M. Shahin, J. Epps, and B. Ahmed, "Automatic classification of lexical stress in English and Arabic languages using deep learning," in *Proc. INTERSPEECH*, San Francisco, CA, USA, Sep. 2016, pp. 175–179.

[70] A. M. C. Sluijter and V. J. van Heuven, "Spectral balance as an acoustic correlate of linguistic stress," *J. Acoust. Soc. Amer.*, vol. 100, no. 4, pp. 2471–2485, Oct. 1996.

[71] D. Neiberg, K. Elenius, and K. Laskowski, "Emotion recognition in spontaneous speech using GMMs," in *Proc. 9th Int. Conf. Spoken Lang. Process. INTERSPEECH (ICSLP)*, Pittsburgh, PA, USA, 2006, pp. 809–812.

[72] F. de Leon and K. Martinez, "Enhancing timbre model using MFCC and its time derivatives for music similarity estimation," in *Proc. 20th Eur. Signal Process. Conf. (EUSIPCO)*, Bucharest, Romania, Aug. 2012, pp. 2005–2009.

[73] P. Tsiakoulis, A. Potamianos, and D. Dimitriadis, "Spectral moment features augmented by low order cepstral coefficients for robust ASR," *IEEE Signal Process. Lett.*, vol. 17, no. 6, pp. 551–554, Jun. 2010.

[74] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digit. Signal Process.*, vol. 10, pp. 19–41, Jan. 2000.

[75] F. Biadsy, "Automatic dialect and accent recognition and its application to speech recognition," Ph.D. dissertation, Dept. Comput. Sci., Graduate School Arts Sci., Columbia Univ., New York, NY, USA, 2011.

[76] M. S. Tiwari, "Text-independent speaker recognition using Gaussian mixture model," *Int. J. Eng. Sci. Manag.*, vol. 2, no. 2, pp. 207–219, Apr./Jun. 2012.

[77] D. O'Shaughnessy, "Acoustic analysis for automatic speech recognition," *Proc. IEEE*, vol. 101, no. 5, pp. 1038–1053, May 2013.

[78] R. B. Lanjewar, S. Mathurkar, and N. Patel, "Implementation and comparison of speech emotion recognition system using Gaussian Mixture Model (GMM) and K-Nearest Neighbor (K-NN) techniques," *Proc. Comput. Sci.*, vol. 49, pp. 50–57, 2015.

[79] C. Zhao, L. Li, D. Wang, and A. Pu, "Local training for PLDA in speaker verification," in *Proc. Conf. Oriental Chapter Int. Committee Coordination Standardization Speech Databases Assessment Techn. (O-COCOSDA)*, Bali, Indonesia, Oct. 2016, pp. 156–160.

NOOR JAMALIAH IBRAHIM received the B.Eng. degree in electrical engineering from the University of Malaysia Pahang (UMP), Pahang, Malaysia, in 2007, and the M.Sc. degree in computer science from the University of Malaya, Kuala Lumpur, Malaysia, in 2010, where she is currently pursuing the Ph.D. degree in speech recognition with the Faculty of Computer Science and Information Technology.

Her current research interests include signal processing, speech recognition, Quranic speech processing, computational technology, and computer-aided learning.

MOHD YAMANI IDNA IDRIS (Member, IEEE) received the B.Eng., M.Sc., and Ph.D. degrees in electrical engineering from the University of Malaya, Kuala Lumpur, Malaysia.

He is currently an Associate Professor with the Department of Computer Systems, Faculty of Computer Science and Information Technology, University of Malaya. He is the author of a book and several articles in reputable journals. His research interests include information security, embedded systems (system on chip and FPGA), image processing and computer vision, digital forensics, surveillance systems, digital signal processing (speech processing and bio-signals), and wireless sensor networks.

M. Y. ZULKIFLI MOHD YUSOFF received the bachelor's degree from the University of Malaya, Kuala Lumpur, Malaysia, the master's degree from The University of Jordan, Amman, Jordan, and the Ph.D. degree from the University of Wales, Lampeter, U.K.

He is currently a Professor with the Department of Al-Quran and Al-Hadith, Academy of Islamic Studies, University of Malaya. He is also the Head of the Centre of Quranic Research (CQR), University of Malaya, and actively engaged in research activities. He is the author of several books and published several articles in reputable journals. His research interests include Quranic exegesis, Quranic studies, and methodology of Quranic exegesis.

ROZIANA RAMLI received the B.Eng. degree in electrical engineering, the M.Sc.Eng. degree in biomedical engineering, and the Ph.D. degree from the University of Malaya, Kuala Lumpur, Malaysia.

She is currently a Senior Lecturer with the Department of Computer Systems, Faculty of Computer Science and Information Technology, University of Malaya. Her current research interests include wireless networks and image/signal processing and analysis.

RAJA JAMILAH RAJA YUSOF (Senior Member, IEEE) received the B.Eng. degree in information system engineering from the Imperial College of Science, Technology and Medicine, London, U.K., and the M.Sc. and Ph.D. degrees in computer science from the University of Malaya, Kuala Lumpur, Malaysia.

She is currently a Senior Lecturer with the Department of Software Engineering, Faculty of Computer Science and Information Technology, University of Malaya. Her main research interests include human–computer interaction and computational algorithms in relation to software engineering while Quranic and Islamic information are valued aspects of the research.