Middle Eastern and North African English Speech Corpus (MENAESC): Automatic Identification of MENA English Accents

Sara Chellali[1], Somaya Al-Maadeed[2], Ouassila Kenai[3], Maamar Ahfir[4], and Walid Hidouci[1]
[1]Laboratory LCSI, Ecole nationale Supérieure d'Informatique, Algeria
[2]Department of Computer Science and Engineering, College of Engineering, Qatar University, Qatar
[3]Laboratory LCPTS, Faculty of Electronics and Computer Sciences, USTHB, Algeria
[4]Department of Computer Science, University Amar Telidji, Algeria

Abstract: *This study aims to explore the English accents in the Arab world. Although there are limited resources for a speech corpus that attempts to automatically identify the degree of accent patterns of an Arabic speaker of English, there is no speech corpus specialized for Arabic speakers of English in the Middle East and North Africa (MENA). To that end, different samples were collected in order to create the linguistic resource that we called Middle Eastern and North African English Speech Corpus (MENAESC). In addition to the "accent approach" applied in the field of automatic language/dialect recognition; we applied also the "macro-accent approach" -by employing Mel-Frequency Cepstral Coefficients (MFCC), Energy and Shifted Delta Cepstra (SDC) features and Gaussian Mixture Model-Universal Background Model (GMM-UBM) classifier- on four accents (Egyptian, Qatari, Syrian, and Tunisian accents) among the eleven accents that were selected based on their high population density in the location where the experiments were carried out. By using the Equal Error Rate percentage (EER%) for the assessment of our system effectiveness in the identification of MENA English accents using the two approaches mentioned above through the employ of the MENAESC, results showed we reached 1.5 to 2%, for "accent approach" and 2 to 3.5% for "macro-accents approach" for identification of MENA English. It also exhibited that the Qatari accent, of the 4 accents included, scored the lowest EER% for all tests performed. Taken together, the system effectiveness is not only affected by the approaches used, but also by the database size MENAESC and its characteristics. Moreover, it is impacted by the proficiency of the Arabic speakers of English and the influence of their mother tongue.*

Keywords: *MENAESC, MFCC+Energy and SDC features, accent, macro-accent, automatic identification.*

Received September 9, 2019; accepted April 8, 2020
https://doi.org/10.34028/iajit/18/1/8

1. Introduction

Speech processing is in increasing demand for many applications, from crime investigation to the simplest daily use (for example, applications on one's mobile). Although recent speech recognition systems have achieved high recognition rates, non-native speech can significantly affect the performance of such systems. Therefore, identification of the non-native speaker's accent can be an auxiliary factor for speech recognition systems, and exploited in determining the identity of the speaker as to his/her country of origin.

The English language has attracted the interest of a great deal of research in automatic identification of dialects/accents (of native or non-native speakers), resulting in a relatively large volume of linguistic resources to serve this research. Some of them were concerned with local and regional dialects of the United Kingdom [9, 10, 15], the United States of America [3, 8, 12], Canada [8], Australia [24] and South Africa [11, 22, 23]. In addition to these resources, the English accents scattered around the world have also received attention, such as European English accents:

German, Spanish, French, Dutch, Italian, Czech, and so on [18, 21, 37, 40], and Asian English accents: Indonesian, Indian, Japanese, Chinese, Malaysian, Thai, Korean, Mongolian, Taiwanese, and others [28, 30, 42, 45, 46]. In this regard, Raab *et al*. [32], in "Non-native speech databases", and Alghamdi *et al*. [2], in "Saudi Accented Arabic Voice Bank", presented an inventory for non-native English databases. In spite of all these efforts, these resources do not cover all the accents that exist in the world, as we noted the almost complete absence of linguistic resources directed to the Arab world.

Despite the significant number of Arabs who speak English, the linguistic resources allocated to this category of speakers are still very few, compared with the English-speaking resources, and research dealing with this category is almost non-existent. To the best of our knowledge, these resources were developed for multilingual automatic speech recognition systems, with a limited number of samples from Arab speakers within this multiple-ethnicity group:

- Center for Spoken Language Understanding (CSLU) (CSLU: foreign accented english release 1.2) consists of spontaneous continuous speech in English by non-native speakers of 22 different mother tongues, among them 112 Arabic utterances (telephone-quality utterances of 20 seconds). However, it is not accessible free of charge (only ten free samples are offered of Arabic speakers) [25].
- The "Speech Accent Archive" is a large set of English accents provided by native and non-native speakers, in which they read the same English paragraph. It was established as a teaching and research tool, containing 172 utterances from Arabic speakers, and it is accessible free of charge [44].
- Australian National Database of Spoken Language (ANDOSL) is a database of Australian English (native and non-native speakers); it contains nine groups of non-native speakers, including Arabic speakers represented by 48 Lebanese speakers. It is not accessible free of charge [24].

The aims of this research are: firstly, to develop a novel English database resource of Arabic speakers of Middle Eastern and North African English (MENA) for systems of automatic accents identification and automatic speech recognition; secondly, to evaluate this resource by developing automatic identification system and testing it.

This paper is organized as follows: section 2 is devoted block diagram of our system with its different modules. In section 3, the linguistic source Middle Eastern and North African English Speech Corpus (MENAESC) that was created specifically for this research purpose and the circumstances of its establishment are described. Section 4 includes the experiments, results and discussion. Finally, the conclusion is in section 5.

2. Automatic Accent Identification System

The block diagram of Figure 1 illustrates our automatic accent identification system:

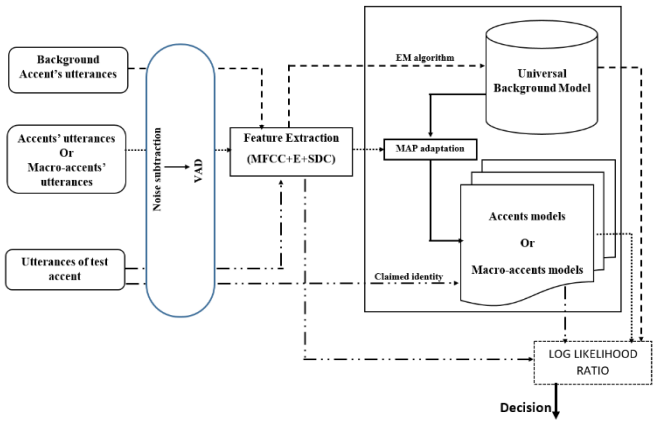

Figure 1. Block diagram of automatic accent identification system.

2.1. Pre-Processing

For improve speech quality, the pre-processing step is important and necessary in speech processing systems. It includes two tasks: noise subtraction and Voice Activity Detection (VAD). Several methods were used for each task. In this work, we base on two methods: Spectral Subtraction Method and Energy-Based VAD, as shown below:

2.1.1. Spectral Subtraction Method

For noise subtraction, we used the spectral subtraction approach proposed by Martin [27] to specify the parameters of the noise estimation algorithm. It is based on tracking the minimum Power Spectral Density (PSD) of noisy speech. Thus for a stationary noise, this approach gives results equivalent to a technique using a robust VAD. Otherwise, if the noise is non-stationary, this approach allows a good follow-up of the evolution of the noise during voice activity, which distinguishes it from techniques based on VAD.

2.1.2. Energy-Based VAD

The mission of the VAD is to determine active speech and inactive speech periods (silence). Several techniques were applied in this domain (energy-based, cepstral coefficients, spectral entropy, a least-square periodicity measure, etc.,). We were interested in the energy-based VAD method due to its simplicity. In this technique, the features extracted from the input signal are compared with a threshold already calculated from silence-only periods, as follows (Algorithm 1):

Algorithm 1: Detection of silence periods.

if ($E_j > k*E_r$), where ($k>1$)) then
  Current frame is Active *speech*
else
  Current frame is Inactive *speech*
end if

E_j is energy of the j^{th} frame, E_r represents the energy of silence frames, while ($k*E_r$) is the "threshold" being used to make the decision.

In this work, we adopted the "dynamic" E-VAD method with an adaptive scaling factor of Sakhnov *et al*. [35, 36]. The main idea is that the threshold level can be estimated by using only the minimums and maximums of speech energy, as:

$$Threshold = k_1 * E_{min} + k_2 * E_{max} \quad (1)$$

Where k_1 and k_2 are the two factors used to interpolate the threshold value to its best performance. It is possible to introduce Equation (2) as a convex combination of a single parameter λ, as:

$$Threshold = (1 - \lambda) * E_{min} + \lambda * E_{max} \quad (2)$$

λ is a scaling factor controlling the estimation process. To achieve a scaling factor that is independent and

resistant to the variable background environment, Equation (3) was suggested:

$$\lambda = \frac{E_{max} - E_{min}}{\mathrm{E}_{max}} \quad (3)$$

It is worth noting that the energy is calculated (by the well-known Root Mean Square Energy (RMSE)), as follows:

$$E_j = \left(\frac{1}{\mathrm{N}} * \sum_{i=(j-1)*N+1}^{j*N} x^2\,(i)\right)^{1/2} \quad (4)$$

Let *x(i)* be the i^{th} sample of speech. If the length of the frame is *N* samples, then the j^{th} frame can be represented as:

$$f_j = \{x(i)\}_{i=(j-1)*N+1}^{j*N} \quad (5)$$

2.2. MFCC and SDC Features Extraction

Because the stages of features extraction and of classification are among the most important steps of the automatic identification process of accents/dialects, many features have been employed, including the Mel-Frequency Cepstral Coefficients (MFCC) [1, 3, 19, 26, 29, 31], Linear Prediction Coefficients (LPC) [41], RelAtive SpecTrAl (RASTA) [15], and Perceptual Linear Prediction (PLP) [13, 15].

The MFCC method is the most popular among the parameter extraction methods; its extraction principle is based on the MEL scale [26]. Indeed, the perception of speech by the human auditory system is based on a frequency scale that is similar to the MEL scale. Recalling that this, scale is linear at low frequencies and logarithmic at high frequencies.

Feature vector extraction for language identification, dialect and accent recognition systems is typically performed by constructing a feature vector at frame time t that consists of cepstra and delta cepstra. However, a previous study [38] showed that improved language identification performance could be obtained by using Shifted Delta Cepstra (SDC) feature vectors created by stacking delta cepstra computed across multiple speech frames.

The SDC features are specified by a set of 4 parameters, *N*, *d*, *P*, and *k*, where *N* is the number of cepstral coefficients computed at each frame, *d* represents the time advance and delay for the delta computation, *k* is the number of blocks whose delta coefficients are concatenated to form the final feature vector, and *P* is the time shift between consecutive blocks. Accordingly, *kN* parameters are used for each SDC feature vector, as compared with *2N* for conventional cepstra and delta-cepstra feature vectors. These features have been chosen because it is well known that the long-time temporal information plays a significant role in capturing language-specific spectral properties [43].

2.3. GMM-UBM Classifier

Numerous classifiers are used for accent recognition, such as Gaussian Mixture Models (GMM) [1, 7, 13, 29], Support Vector Machine (SVM) [4, 31, 39], Artificial Neural Network (ANN), Deep Neural Networks (DNN) and Recurrent Neural Network (RNN) [5, 17, 20, 26], Hidden Markov Model (HMM) [3, 16, 19, 24, 39, 41], K-Nearest Neighbor (KNN) [26], to achieve high identification accuracy. Nevertheless, there is disagreement in the field as to the best of the above methods, because the effectiveness of the identification system is influenced by many other factors, such as the database, that are often closely inter-related.

The Gaussian Mixture Model-Universal Background Model (GMM-UBM) has been successfully applied in identifying speakers, language and dialects [13, 15, 33]. The creation of an accent identification system based on GMM-UBM must pass through these stages:

1. The generation of the UBM model and accents models by using utterances from all accents. An independent UBM model is generated with a large GMM, which is built based on the Estimation Maximization (EM) algorithm. Then, for each accent, a GMM is derived via maximum a posteriori (MAP) estimation from the UBM and available training data for each accent (a specific model is obtained for each accent through the adaptation of the UBM parameters, which couples the MAP-adapted model and the UBM).
2. In the test phase, the matching score depends on the accent target model (M_{tar}) and background model (M_{UBM}) via the Log Likelihood Ratio (LLR):

$$LLR = log\ P(X/M_{tar}) - logP(X/M_{UBM}) \quad (6)$$

3. The GMM with the highest likelihood (LLR) gives the accent classification, or the obtained LLR should be compared to a decision threshold θ to accept (or reject) the claimant accent.

3. Middle Eastern and North African English Speech Corpus (MENAESC)

We became aware that free or paid English language databases in the MENA region are very limited, so we created a new database in two phases: the first version contained 749 utterances from 19 speakers [6], and was subsequently extended to the current (second) version with 3,224 utterances of 83 speakers distributed across 11 local accents.

3.1. The MENA and its Arabic Dialects

The MENA, or the Arab world, is divided into 22 countries. It has an area of 14 million km^2, equivalent to 10.2% of the world's area, and contains about 6% of

the world's population. Although Modern Standard Arabic (MSA) is the official language of the Arab countries, it is used only in academic and administrative transactions; the predominant language in transactions and daily dialogues is Modern Colloquial Arabic (MCA), which is divided into six basic macro-dialects (Figure 2) [14]:

- Gulf Arabic: the dialects of Bahrain, Kuwait, Oman, Qatar, Saudi Arabia and United Arab Emirates (UAE).
- Iraqi Arabic: classified as a sub-dialect of Gulf Arabic.
- Egyptian Arabic: the dialects of Egypt and Sudan.
- Levantine Arabic: the dialects of Jordan, Lebanon, Palestine and Syria.
- Maghreb Arabic: the dialects of North Africa (Algeria, Mauritania, Morocco, Libya and Tunisia).
- Yemenite Arabic.

The notion of the macro-accents introduced in this present study is a concept that we launched to express a group of the dialects geographically adjacent. Besides these dialects belong to the same language, they also share some linguistic characteristics. In turn each dialect contains various micro-dialects.

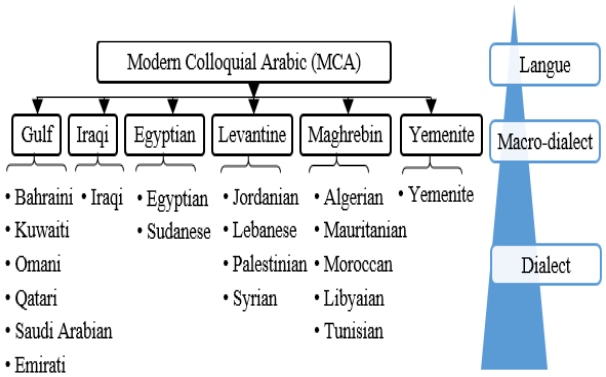

Figure 2. MCA dialectal architecture.

3.2. Fieldwork Site and Participants

Audio samples were collected at Qatar University in Doha, in different locations on the campus. Qatar is on a peninsula in the Arabian Gulf located east of the Arabian Peninsula, with the following ethnic distribution: 40% Arabs, 18% Indians, 18% Pakistanis, 10% Iranians, and 14% other.

During sample collection, we took into consideration the varying quality of records: with and without background sounds, in sound-buffered rooms, and outdoors (to introduce reverberation and echo elements in the phonograms, as well as to obtain a natural environment containing various types of chaos, such as sounds of nature, voices of other people, etc.,).

Eighty-three speakers from 11 different Arab nationalities were involved in the audio recording process (Table 1). The audio recording process included people from different educational levels who were available, including teachers, employees, students, and researchers (43 females and 40 males), aged from 21 to 55 years. The speakers' English learning background ranges from beginner to advanced level. This choice was made to cover maximum population variability. After beginning with these MENA English accents, the database would then be extended to include the remaining MENA English accents.

Table 1. Distribution of MENAESC by macro-accents/accents and speakers/samples.

Macro-Accents	Accents	Number of Speakers	Number of Samples
Egyptian	Egyptian	13	516
	Sudanese	8	268
Gulf	Qatari	18	689
	Saudi Arabian	1	30
Iraqi	Iraqi	10	339
Levantine	Jordanian	2	99
	Lebanese	2	98
	Palestinian	8	358
	Syrian	10	309
Maghrebin	Algerian	3	156
	Tunisian	8	362
Total		**83**	**3,224**

3.3. Data Collection

Voice recording was performed in several sessions using a Sony Dictaphone (Stereo Integrated Circuits (IC[1]) Recorder, Integrated Circuits and Devices (ICD[2]-UX560F). We asked each speaker to repeat a reading five times in one session, the first ten numbers, twelve isolated words divided into five groups, five short sentences, and a paragraph of 69 words. It was not a spontaneous reading but rather a prepared reading from each speaker. We tried to take into account some linguistic characteristics in the choice of text (digits, isolated words, short sentences and paragraphs) [6, 44], such as the phonetic differences between Arabic and English. The length of the recordings ranged from two to six minutes each (some speakers did not adhere strictly to the number requested).

Table 2. Distribution of the MENAESC samples by category/duration.

Type of Sample	Number of Samples	Duration (s.ms)		Total Duration
		Min	Max	
Number	338	02.443	10.902	00:34:44.528
Isolated words	1274	01.789	10.198	01:26:10.178
Short sentences	1300	00:665	04.476	00:35:04.379
Paragraph	312	16.476	30.000	02:20:01.727
Total	**3224**	**00.665**	**30.000**	**04:56:00.812**

The recordings were processed and edited into sub-recordings of a short length (less than thirty seconds). Every sub-recording contained one utterance: the ten numbers, one group of isolated words, one sentence or the paragraph. After editing the recordings into shorter files, every file was coded with a nine-digit code

[1]Integrated Circuit
[2]Integrated Circuits and Devices

corresponding to information about the country, city, number of speakers and number of utterances [6].

This corpus consists of utterances pronounced by Arab speakers. There are 3,224 utterances (each utterance is a wave type audio file in stereo, with a sampling rate equal to 44.1 kHz and a quantization level of 16 bits) from native speakers of 11 Arabic dialects divided into 5 macro-dialects. This represents almost 4 hours and 56 minutes of speech recording collected (Table 2).

4. Experiments and Results

All Training and testing data used were from the MENAESC database. This corpus consists of utterances pronounced by Arab speakers. There are 3,224 utterances (stereo, F_s=44,100 Hz, .wav) from native speakers of 11 Arabic accents divided on 5 Macro-accents.

4.1. Protocol of Experiments

For the testing phase, the Egyptian, Qatari, Syrian and Tunisian accents were selected. The choice of these 4 accents of test was based on their high population density in the place where the recordings were carried out. In addition, our database does not contain the Yemeni accent, which led to its exclusion from our experiments, while the Iraqi dialect is classified as a sub-accent of the Arabian Gulf.

From these 4 accents, samples used in the testing phase were randomly selected. The tests of data samples for Validation set (Valset) and Evaluation set (Evaset) were conducted on Egyptian, Qatari, Syrian, and Tunisian accents, as shown in Table 3.

Table 3. Distribution of testing data samples.

Accents	Number of testing data samples	
	Valset	Evaset
Egyptian	20	50
Qatari	20	50
Syrian	20	50
Tunisian	20	50
Total	**80**	**200**

The Valset tests included 80 samples, while the Evaset tests included 200 samples. The samples in the Valset are known and included in the training data and the samples in the Evaset are completely unknown and not included in the training data.

As outlined below, a series of various experiments was carried out based on the number of coefficients to extract and the number of GMM to use, in order to achieve optimal performance of our identification system.

For the number of features, we tested for 12 MFCC, 12 MFCC+Energy, 13 MFCC (includes the 0^{th} coefficient) and 13 MFCC+Energy; and we tested for 24, 26, 36, 39, 56 and 64 SDC. In addition, for the number of GMM, we tested for 32, 64, 128, 256, 512, 1024, 2048, and 4096 GMM.

In this study, the characteristics of the SDC are derived from the MFCC by setting to 13-2-3-3. This is the custom setting validated by many tests performed on our database. The dimension of the SDC entities is 39 with 12 MFCC and E (Energy) to produce a 52-dimensional feature vector.

We used the Microsoft Research (MSR) Identity Toolbox, VOICEBOX (Speech Processing Toolbox) [34], and MATLAB for development.

Two principal experiments were conducted, depending on the training data (Table 4):

- The first experiment "Accent approach": The Egyptian, Qatari, Syrian and Tunisian accents chosen among the 11 accents to create the 4 GMM models.

Table 4. Distribution of samples/accents and samples/macro-accents

1st Training Dataset		2nd Training Dataset	
Accents	**Number of Samples**	**Macro-accents**	**Number of Samples**
Egyptian	436	Egyptian	704
Sudanese	268		
Qatari	609	Gulf +	978
Saudi Arabian	30		
Iraqi	339		
Jordanian	99	Levantine	784
Lebanese	98		
Palestinian	358		
Syrian	229		
Algerian	156	Maghrebin	438
Tunisian	282		
Total	**2,904**	**Total**	**2,904**

- The second experiment "Macro-accent approach": The Egyptian, Gulf[+3], Levantine and Maghrebin macro-accents were selected to create the 4 GMM models.

The eleven accents (Algerian, Egyptian, Iraqi, Jordanian, Lebanon, Palestinian, Qatari, Saudi Arabian, Syrian, Sudanese, and Tunisian) that exist in the MENAESC were selected to create the UBM in two experiments. It included 2904 samples (audio files).

4.2. Results and Discussion

Recalling that we have created a UBM for all accents (11 accents) and GMM model for each accent or macro-accent previously mentioned.

From the different tests performed for both experiments, we see that our identification system achieved its best results with GMM = 1024 and 2048, 12 MFCC+E and 12 MFCC+E+39 SDC for Valset and Evaset.

The Tables 5 and 6, respectively, show results of the first and second experiments by employing the percentage (EER%). The EER value is the intersection between the False Rejection Rate (FRR) and the False Acceptance Rate (FAR).

For GMM=2,048: the best results EER%=0% for

[3]Gulf +=Gulf + Iraqi

Valset and, against the EER%=2% for Evaset, for the first experiment "Accent approach". However, for the second experiment "Macro-accent approach": the best results EER%=00% for Valset, against the EER%=3.5% for Evaset.

Table 5. Results EER% performances of the first experiment.

Features / GMM-UBM	MFCC+E		MFCC+E+SDC	
	Valset	Evaset	Valset	Evaset
GMM=1024	0.8333%	1.5%	0%	3%
GMM=2048	**0.4167%**	**1.5%**	**0%**	**2%**

Table 6. Results EER% performances of the second experiment.

Features / GMM-UBM	MFCC+E		MFCC+E+SDC	
	Valset	Evaset	Valset	Evaset
GMM=1024	1.25%	3.5%	0%	4.5%
GMM=2048	**0%**	**2%**	**0%**	**3.5%**

These results show the full efficiency of our system, in the case of known utterances (Valset) is tested, and among those found in the training base, despite the fact that the statements are completely different the performances EER%= 0% when using MFCC+E+SDC for GMM=1,024 and 2,048. Also they show our system is stable in case of unknown utterances (Evaset) is tested when using MFCC+E for both GMM numbers the performances EER%=1.5%, and slight degradation of EER%=2% when using MFCC+E+SDC for GMM=2,048.

We also noticed that there is no significant difference in results for our identification system, when using either distributions of MENAESC (macro-accents or accents), which confirms that the representation in macro-accents is not better and not worst if compared to the results for Evaset in the two experiments.

As with all databases in the experiments, the optimal values are for GMM=2048 and MFCC+E+SDC for Vadset and for the same number GMM with 12 MFCC+E for Evaset.

Given the details of the EER% for each accent (Table 7), we find that:

Table 7. Comparison of EER% performances between identification of MENA English accents using the accent and Macro-accent approaches.

	Evaset		Average EER%
	Accents approach	**Macro-accents approach**	
Egyptian	**0.6667%**	1.5%	1,08
Qatari	**1.3333%**	**0.5%**	**0.91**
Syrian	2.6667%	10%	6,33
Tunisian	2.0000%	**0.5%**	1.25

The Syrian accents (in varying order, according to the testing experiments) register a relatively high rate, where we note that the ERR% for the Egyptian, Tunisian and Qatari accent is the best in all tests performed in first experience. However, in second experiment, we note that the performance of EER% is better and this confirms that the Qatari and Tunisian accents are more distinctive than for the other accents

In the first experiments, for the Egyptian accent achieves the lowest of EER% = 0.6667 %, followed by the Qatari, Tunisian and Syrian accents, respectively. In the second experiment, the Qatari and Tunisian accents achieve the lowest of EER% = 0.5% followed by the Egyptian and Syrian accents, respectively; while the Syrian accent registers every time the relatively high rate. By calculating the average EER for all the accents we find for the Qatari accent achieves the lowest of average EER% = 0.91%, so it's the best if we compare it to the other average EER for the other accents.

Table 8 reveals that misclassifications are distributed across different accents with convergent values, except for the Qatari and Tunisian accents, which are distinct for macro-accents experience where the identification is 100% perfect then Egyptian accent. This is somewhat predictable, because the speakers share the same mother tongue, Arabic, along with the influence of the media on the spread of some Arabic dialects, the greater employment opportunities offered by some countries, and mixed marriages between them.

We present in this histogram (Figure 3) the performances of the EER% the results of all the tests carried out for two experiments (the accents approach and macro-accent approach); we notice in the case where we have tested our system for the 4 accents the EER% reaches the best scores for the accent Egyptian and Qatari then Tunisian and Syrian, respecly. On the other hand, for the 4 macro-accents the ERR% reaches the highest scores for Qatari and Tunisian then Egyptian and last Syrian. In order to normalize the scores, we calculated the average of EER% for each accent for both experiences, and we find the best one is for the Qatari accent and the lowest for Syrian. All these results confirm that our system can more easily identify the Qatari accent due to their native language compared to other accents.

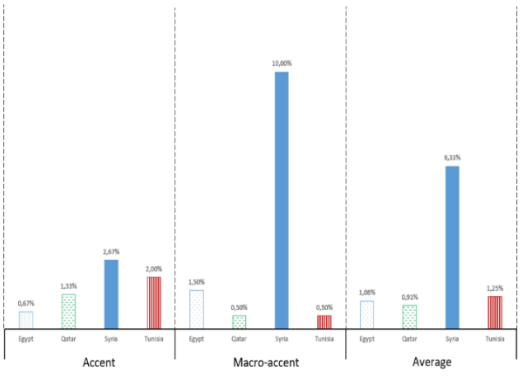

Figure 3. Histogram the EER% performances for identification of MENA English accents using the accent and macro-accent approaches.

5. Conclusions

The objective of this study is to present a linguistic

source for non-native English accents, especially for Arabic speakers of English from the MENA, by using a system of automatic identification of accents. The novel MENAESC has been created and evaluated. The validation and the evaluation tests on training data were carried out on the Egyptian, Qatari, Syrian and Tunisian accents; where we used the eleven accents that exist in the MENAESC in order to create the UBM; and for adapt GMM models we used four accents in the first experiment and four macro-accents in the second experiment. Results of the experiments were 1.5 and 3.5% for the evaluation's tests, and between 0 and 0.4167% for the validation tests using 2,048 GMM and 52 features (MFCC+E+SDC). These results reflect a relatively good performance of the automatic identification of MENA English accents, considering that this study is just the initial attempt to classify the English accents of native Arabic speakers. It may be attributed to the nature of the MENA English accents/macro-accents, particularly as they include distinctive characteristics among their other shared characteristics. This pattern plays a role in the difficulty of identifying and discriminating between the accents/macro-accents, especially those that are geographically adjacent. In addition, the small size of MENAESC and the level of the speakers' proficiency in English made the identification process even more complex. Inevitably, the influence of other factors on the identification system must be taken into account, such as methods and approaches used as well as the recording conditions (material, environment ...).

Acknowledgment

This paper was made possible by a QUCP award [QUCP-CENG-CSE-15-16-1] from the Qatar University. The statements made herein are the sole responsibility of its authors.

Table 8. Confusion matrices results obtained for identification of MENA English accents of Evaset (for GMM =2,048, 12 MFCC+ E+39 SDC).

	Accents approach					**Macro-accents approach**			
	Egyptian	**Qatari**	**Syrian**	**Tunisian**		**Egyptian**	**Gulf**	**Levantine**	**Maghrebin**
Egyptian	**49**	0	0	1	**Egyptian**	**49**	0	0	1
Qatari	1	**49**	0	0	**Qatari**	0	**50**	0	0
Syrian	0	2	**48**	0	**Syrian**	0	2	**45**	3
Tunisian	0	0	1	**49**	**Tunisian**	0	0	0	**50**

References

[1] Abed A. and Guerti M., "HMM/GMM Classification for Articulation Disorder Correction among Algerian Children," *The International Arab Journal of Information Technology*, vol. 13, no. 4, pp. 449-455, 2016.

[2] Alghamdi M., Alhargan F., Alkanhal M., Alkhairy A., Eldesouki M., and Alenazi A., "Saudi Accented Arabic Voice Bank," *Journal of King Saud University-Computer and Information Sciences*, vol. 20, pp. 45-62, 2008.

[3] Arslan L. and Hansen J., "Language Accent Classification in American English," *Speech Communication*, vol. 18, no. 4, pp. 353-367, 1996.

[4] Bahari M., Saeidi R., Van hamme H., and Van Leeuwen D., "Accent Recognition Using I-Vector, Gaussian Mean Super Vector and Gaussian Posterior Probability Super Vector for Spontaneous Telephone Speech," *in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, pp. 7344-7348, 2013.

[5] Blackburn C., Vonwiller J., and King R., "Automatic Accent Classification Using Artificial Neural Networks," *in Proceedings of 3rd European Conference on Speech Communication and Technology*, Berlin pp. 1241-1244, 1993.

[6] Chellali S., Al-Maadeed S., Kenai O., Ahfir M., and Hidouci W., "Construction of Audio Corpus of Nonnative English Dialects-Arabs Speakers-," *in Proceedings of the 4th International Conference on Artificial Intelligence and Pattern Recognition*, Poland, pp. 98-102, 2017.

[7] Choueiter G., Zweig G., and Nguyen P., "An Empirical Study of Automatic Accent Classification," *in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, pp. 4265-4268, 2008.

[8] Cieri C., Miller D., and Walker K., "The Fisher corpus: A Resource for the Next Generations of Speech-to-Text," *in Proceedings of the 4th International Conference on Language Resources and Evaluation*, Lisbon, pp. 69-71, 2004.

[9] D'Arcy S., Russell M., Browning S., and Tomlinson M., "The Accents of the British Isles (ABI) Corpus," *in Proceeding of Modélisations Pour Identification des Langues*, Paris, pp. 115-119, 2005.

[10] De Marco A. and Cox S., "Iterative Classification of Regional British Accents in I-Vector Space," *in Proceedings of Symposium on Machine Learning in Speech and Language Processing MLSLP*, USA, pp. 1-4, 2012.

[11] De Wet F., Louwa P., and Niesler T., "Human and automatic Accent Identification of Nguni and Sotho Black South African English," *South*

African Journal of Science, vol. 103, no. 3, pp. 159-164, 2007.

[12] Garofolo J., Lamel L., Fisher W., Fiscus J., and Pallett D., "David S DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CD-ROM, NIST Speech disc 1-1.1," NASA STI/Recon Technical Report, 1993.

[13] Ge Z., "Improved Accent Classification Combining Phonetic Vowels with Acoustic Features," *in Proceedings of 8th International Congress on Image and Signal Processin*, Shenyang, pp. 1204-1209, 2015.

[14] Habash N., "Introduction to Arabic Natural Language Processing," *Synthesis Lectures on Human Language Technologies*, vol. 3, no. 1, pp.1-87, 2010.

[15] Hanani A., Russell M., and Carey M., "Human and Computer Recognition of Regional Accents and Ethnic Groups from British English Speech," *Computer Speech and Language*, vol. 27, no. 1, pp. 59-74, 2013.

[16] Hansen J. and Arslan L., "Foreign Accent Classification Using Source Generator Based Prosodic Features," *in Proceedings of International Conference Acoustic, Speech Signal Process*, Detroit, pp. 836-839, 1995.

[17] Hautamaki V., Siniscalchi S., Behravan H., Salerno V., and Kukanov I., "Boosting Universal Speech Attributes Classification with Deep Neural Network for Foreign Accent Characterization," *in Proceedings of 16th Annual Conference of the International Speech Communication Association*, Dresden, pp. 408-412, 2015.

[18] Heuvel H., Choukri K., Gollan C., Moreno A., and Mostefa D., "TC-STAR: New Language Resources for ASR and SLT Purposes," *in Proceedings of the 5th International Conference on Language Resources and Evaluation*, Genoa, pp. 2570-2573, 2006.

[19] Humphries J., Woodland P., and Pearce D., "Using Accent-Specific Pronunciation Modelling for Robust Speech Recognition," *in Proceedings of the 4th International Conference on Spoken Language*, Philadelphia, pp. 2324-2327, 1996.

[20] Jiao Y., Tu M., Berisha V., and Liss J., "Accent Identification by Combining Deep Neural Networks and Recurrent Neural Networks Trained on Long and Short-Term Features," *in Proceedings of Interspeech Native Language Sub-Challenge*, San Francisco, pp. 2388-2392, 2016.

[21] Jurafsky D., Wooters C., Tajchman G., Segal J., Stolcke A., Fosler E., and Morgan N., "The Berkeley Restaurant Project," *in Proceedings of the International Conference on Spoken Language Processing*, Yokohama, pp. 2139-2142, 1994.

[22] Kamper H. and Niesler T., "Multi-Accent Speech Recognition of Afrikaans, Black and White Varieties of South African English," *in Proceedings of 12th Annual Conference of the International Speech Communication Association*, Florence, pp. 3189-3192, 2011.

[23] Kamper H., Mukanya F., and Niesler T., "Acoustic Modelling of English Accented and Afrikaans Accented South African English," *in Proceedings of PRASA*, Stellenbosch, pp. 117-122, 2010.

[24] Kumpf K. and King R., "Automatic Accent Classification of Foreign Accented Australian English Speech," *in Proceeding of 4th International Conference on Spoken Language Processing*, Philadelphia, pp. 1740-1743, 1996.

[25] Lander T., "CSLU: Foreign Accented English Release 1.2," Linguistic Data Consortium, Philadelphia: Linguistic Data Consortium, 2007.

[26] Ma Z. and Fokoué E., "A Comparison of Classifiers in Performing Speaker Accent Recognition Using Mfccs," *Open Journal of Statistics*, vol. 4, no. 4, pp. 258-266, 2014.

[27] Martin R., "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 5, pp. 504-512, 2001.

[28] Minematsu N., Tomiyama Y., Yoshimoto K., Shimizu K., Nakagawa S., Dantsuji M., and Makino S., "Development of English Speech Database Read by Japanese to Support CALL Research," *Intelligent Control and Automation*, pp. 577-560, 2004.

[29] Nguyen P., Tran D., Huang X., and Sharma D., "Australian Accent-Based Speaker Classification," *in Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*, Phuket, pp. 416-419, 2010.

[30] Patel I., Kulkarni R., and Yarravarapu S., "Automatic Non-Native Dialect and Accent Voice Detection of South Indian English," *Advances in Image and Video Processing*, vol. 5, no. 1, pp. 39-48, 2017.

[31] Pedersen C. and Diederich J., "Accent Classification Using Support Vector Machines," *in Proceedings of the 6th IEEE/ACIS International Conference on Computer and Information Science*, Melbourne, pp. 444-449, 2007.

[32] Raab M., Gruhn R., and Noeth E., "Non-Native Speech Databases," *in Proceedings of Automatic Speech Recognition and Understandin*, Kyoto, pp. 413-418, 2007.

[33] Reynolds D., Quatieri T., and Dunn R., "Speaker Verification Using Adapted Gaussian Mixture Models," *Digital Signal Processing*, vol. 10, no.

1-3, pp. 19-41, 2000.

[34] Sadjadi S., Slaney M., and Heck L., "MSR Identity Toolbox v1.0: A MATLAB Toolbox for Speaker-Recognition Research," Microsoft Research Technical Report, 2013.

[35] Sakhnov K., Verteletskay E., and Simak B., "Approach for Energy-Based Voice Detector with Adaptive Scaling Factor," *IAENG International Journal of Computer Science*, vol. 36, no. 4, 2009.

[36] Sakhnov K., Verteletskaya E., and Šimák B., "Dynamical Energy-Based Speech/Silence Detector for Speech Enhancement Applications," *in Proceedings of the World Congress on Engineering*, London, 2009.

[37] Segura J., Ehrette T., Potamianos A., Fohr D., Illina I., Breton P., Clot V., Gemello R., Matassoni M., and Maragos P., "The HIWIRE Database, A Noisy and Non-Native English Speech Corpus for Cockpit Communication," http://www.hiwire.org/, 2007.

[38] Singer E., Torres-Carrasquillo P., Gleason T., Campbell W., and Reynolds D., "Acoustic, Phonetic, Discriminative Approaches to Automatic Language Identification," *in Proceedings of 8th European Conference on Speech Communication and Technology*, Geneva, pp. 1345-1348, 2003.

[39] Tang H. and Ghorbani A., "Accent Classification Using Support Vector Machine and Hidden Markov Model," *in Proceedings of Conference of the Canadian Society for Computational Studies of Intelligence*, Halifax, pp. 629-631, 2003.

[40] Teixeira C., Trancoso I., and Serralheiro A., "Recognition of Non-Native Accents," *in Proceedings of 5th European Conference on Speech Communication and Technology*, Rhodes, pp. 2375-2378, 1997.

[41] Teixeira C., Trancoso I., and Serralheiro A., "Accent Identification," *in Proceedings of 4th International Conference on Spoken Language Processing*, Philadelphia, pp.1784-1787, 1996.

[42] Visceglia T., Tseng C., Kondo M., Meng H., and Sagisaka Y., "Phonetic Aspects of Content Design in AESOP (Asian English Speech Corpus Project)," *in Proceedings of Oriental COCOSDA International Conference on Speech Database and Assessments*, Urumqi, pp. 60-65, 2009.

[43] Wang H., Leung C., Lee T., Ma B., and Li H., "Shifted-Delta MLP Features for Spoken Language Recognition," *IEEE Signal Processing Letters*, vol. 20, no. 1, pp. 15-18, 2013.

[44] Weinberger S., Speech Accent Archive, http://accent.gmu.edu, Last Visited, 2019.

[45] Yusnita M., Paulraj M., Yaacob S., Abu Bakar S., and Shahriman A., "Malaysian English Accents Identification Using LPC and Formant Analysis," *in Proceedings of IEEE International Conference on Control System, Computing and Engineering*, Penang, pp. 472-476, 2011.

[46] Yusnita M., Paulraj M, Sazali Y., Yusuf R., and Shahriman A., "Analysis of Accent-Sensitive Words in Multi-Resolution Mel-Frequency Cepstral Coefficients for Classification of Accents in Malaysian English," *International Journal of Automotive and Mechanical Engineering*, vol. 7, pp. 1053-1073, 2013.

Sara Chellali is a Ph.D. student at "Ecole nationale Supérieure d'Informatique (ESI, ex INI) ", Algiers, Algeria. She received the Magister degree in Computer Sciences and the Master degree in Didactics of French as a foreign language from University of Amar Telidji, Laghouat, Algeria, and Engineer degree in Computer Systems from ESI. She is currently working as teacher-researcher in the "École Normale Supérieure de Laghouat ENSL", Laghouat, Algeria. Her research is in language processing with particular emphasis on identification of accent/dialect, speech processing, deep learning, machine learnnig, pattern recognition and didactic of sciences (Mathematics).

Somaya Al-Maadeed is a professor at Computer Science and Engineering Department at Qatar University. She received the Ph.D. degree in computer science from Nottingham, U.K., in 2004. She supervised students through research projects related to pattern recognition and Arabic recognition. She is currently the Head of the Computer Science Department, Qatar University. She is also the Coordinator of the Computer Vision Research Group, Qatar University. She enjoys excellent collaboration with national and international institutions, and industry. She is a principal investigator of several funded research projects generating approximately five million dollars in the last years. She published extensively in computer vision and pattern recognition and delivered workshops on teaching programming for undergraduate students. She attended workshops related to higher education strategy, assessment methods, and interactive teaching. In 2015, she was elected as the IEEE Chair of the Qatar Section. She and her team were the recipient of the best performance at ICDAR 2011 and ICDAR 2015 signature verification.

Ouassila Kenai is Ph.D. student in speech communication in USTHB, Algiers, Algeria. She has got Magister degree in automatic speech processing from the Scientific and Technical Research Center for the Development of the Arabic Language CRSTDLA, Algeria and Engineer degree in communication (Electronics) from USTHB, Algeria. She is currently teacher at the institute of trades performing arts and audiovisual ISMAS, Algiers, Algeria. She also works as a teacher and consultant in the audiovisual field in several state and private establishments. Her research interests include speaker recognition -where she presented a new architecture based VAD for speaker diarization/detection systems (it was the subject of a published article)-, artificial intelligent, bioinformatics, speech and language processing, and forensic recognition (She published several conference papers on it).

Maamar Ahfir received his "Ingeniorat" in Electronics and "Magister" in Optoelectronics, both from the University of Blida (Algeria), respectively in 1990 and 1997. He holds the E-science Doctorate degree in Electronics since 2008 from the "Ecole Nationale Polytechnique (ENP)" of Algiers (Algeria). He was Lecturer in the University of Laghouat (Algeria) from 1997 to 2019 and Head of the Informatics Department of the Technical College of Jizane (Saudi Arabia) from 2001 to 2002. He is currently Associate Professor at the University of Médéa (Algeria) since 2019 and Visiting Researcher to Applied DSP and VLSI Systems Laboratory of the University of Westminster, London, UK, since 2004.His areas of interest include room acoustics, speech and human heart sounds (Phonocardiogram) processing.

Walid Hidouci is a professor in computer science at "Ecole nationale Superieure d'Informatique: ESI" in Algiers. He leads the "Advanced Database Systems" team in the LCSI research laboratory. His main topics of interests are: database systems, data structures, artificial intelligence, operating systems and parallel programming.