Helpful Statistics in Recognizing Basic Arabic Phonemes

Mohamed O.M. Khelifa
TES Research Team
ENSIAS School of Engineering
Mohammed V University in RABAT
Rabat, Morocco

Yahya O.M. ElHadj
Doha Institute, Doha, Qatar
SAMoVA Research Team, IRIT
Paul Sabatier University
Toulouse, France

Yousfi Abdellah
FSJES-souissi
Mohammed V University in RABAT
Rabat, Morocco

Mostafa Belkasmi
TES Research Team
ENSIAS School of Engineering
Mohammed V University in RABAT
Rabat, Morocco

***Abstract*—The recognition of continuous speech is one of the main challenges in the building of automatic speech recognition (ASR) systems, especially when it comes to phonetically complex languages such as Arabic. An ASR system seems to be actually in a blocked alley. Nearly all solutions follow the same general model. The previous research focused on enhancing its performance by incorporating supplementary features. This paper is part of ongoing research efforts aimed at developing a high-performance Arabic speech recognition system for learning and teaching purposes. It investigates a statistical analysis of certain distinctive features of the basic Arabic phonemes which seems helpful in enhancing the performance of a baseline HMM-based ASR system. The statistics are collected using a particular Arabic speech database, which involves ten different male speakers and more than eight hours of speech which covers all Arabic phonemes. In HMM modeling framework, the statistics provided are helpful in establishing the appropriate number of HMM states for each phoneme and they can also be utilized as an initial condition for the EM estimation procedure, which generally, accelerates the estimation process and, thus, improves the performance of the system. The obtained findings are presented and possible applications of automatic speech recognition and speaker identification systems are also suggested.**

Keywords*—automatic speech recognition (ASR); speech recognizer; phonemes recognition; speech database; hidden Markova models (HMMs)

I. INTRODUCTION

The most communal way for humans to communicate is through sounds made during speech operation. Thoughts and ideas are exchanged via speech. One person speaks and the other receives the message by means of their ears. Automatic speech recognition (ASR) is the process by which a computer is capable of recognizing and acting upon spoken language or utterances using particular algorithms [1-5]. It is a branch of artificial intelligence (AI) and is related to various areas of knowledge, including informatics, linguistics, acoustics, and pattern recognition. An ordinary ASR system consists of a microphone unit, speech recognition engine, computer, and a certain form of audio/visual/action output. The Applications of an ASR system can be classified into two main areas. One is dictation, and the other is human-computer dialogue applications. In the dictation area, the broadcast news dictation technology has been incorporated into information extraction and retrieval technology, and many application systems such as retrieval systems and automatic voice document indexing. In the human-computer interaction area, a variety of experimental systems for information retrieval through spoken dialogue were investigated. A common ASR application is the automated conversion of speech into written text, which has the capability to increase output effectiveness and enhance access to diverse computer applications such as word processing, email, remote control, using phones, language identification, speaker identification, and archiving and language acquisition.

By using speech as input, ASR applications reduces the more traditional manual input techniques via keyboards and mousses, making it helpful as an alternative input technique for people with disabilities. ASR performance may be affected by various factors, including the quality of the inputted speech, the technology design, the surrounding environment and speaker characteristics.

In spite of the remarkable advances in signal processing, computational architectures, algorithms and hardware, ASR systems is still a topic of an active research and ideal systems are still far from reached [6]. Thus, the most important research issues should be attacked in order to advance to the ultimate goal of fluent speech recognition.

In speech recognition, it is uncomplicated to recognize isolated words but the main challenge is to recognize continuous speech. There are two parts for any ASR system: the language model and the acoustic model. The language model indicates the status of word sequences to be recognized: are they common or rare? Thereby, the acoustic model is used to model the sounds we produce when we speak. For a small vocabulary, it's easy to model the acoustics of individual words. As vocabulary size grows, it becomes impractical to

record sufficient spoken examples of all words and so we need to model acoustics at a lower level. The state-of-the-art ASR systems do not rely on the whole words in both training and decoding process due to the enormous quantity of words that may exist in a speech corpus in addition to the necessity to have sufficient spoken examples for each word. Contrariwise, a successful ASR system uses smaller parts of words or sub-word units of words that are commonly designed by phoneticians or expert in linguistics. This set of sub-word units is referred to as phonemes.

Most of the current successful ASR systems are based on hidden Markov models (HMM) in which each phoneme is modeled by a set of HMM states. A 3 emitting states with left-to-right HMM topology are commonly used for each phoneme independent of its length. Thus, the question that arises is whether this number of states is sufficient for certain phonemes or is it greater or fewer than what is needed? One of the main matters in ASR system is to determine the number of HMM states that reflects the correct length of each phoneme occurrence in a speech corpus.

Despite the sizable utilization of speech recognition technologies in foreign languages likes English and French, Arabic the rarity of mature ASR-based applications, especially for language teaching and learning. One renowned application of Arabic Speech Recognition is the teaching of Classical Arabic (CA) sound system. Although classical Arabic is not utilized in everyday communication, it is required for learning the Holy Quran (The Muslim Holy Book) and the old Arabic poetry heritage. Moreover, it can open the door for various sorts of Islamic applications.

The present paper is part of ongoing research efforts aiming to develop a high-performance Arabic speech recognition system for learning and teaching purposes. First stages of these efforts were dedicated to the development of particular Arabic speech database including ten different speakers and more than eight hours of speech collected from recitations of the Holy Quran in which all Arabic phonemes are included. Speech signals of this speech database were manually and accurately segmented and labeled on three levels: word, phoneme, and allophone. Next, two baselines HMM-based recognizers were built to validate the speech segmentation on both phoneme and allophone levels and also to examine the intended recognition accuracy in both recognizers.

This current stage investigates a statistical analysis of certain distinctive features in Arabic phonemes in order to incorporate them later into the speech recognition process for the aim of improving the performance of our baseline HMM-based recognizers. The distinctive features which have been investigated in this work are phoneme durations, mean durations of phonemes, median of the duration for each basic phoneme, median of the durations, frequency and probability occurrences for each basic phoneme. Analysis and interpretations were performed to determine which of these distinctive features can significantly enhance systems performance. In HMM modeling framework, the statistics provided can be helpful in establishing the appropriate number of HMM states for each phoneme which generally increases the speed and recognition accuracy. The phonemes statistics can also be utilized as an initial condition for the Expectation-Maximization estimation procedure and hence accelerates the estimation process, or it can be utilized as a wanted model itself. Also, the probability of the neighboring two phoneme clusters is helpful information which is not yet integrated in the adjustment of speech characteristics of possible words from a dictionary.

The rest of the article is organized as follows: section 2 summarizes our research efforts accomplished towards the ultimate goal. Section 3 describes the motivation of the presented work. Section 4 introduces a brief overview of the previously developed speech database. In Section 5 we present the methodology used for statistics extraction. Section 6 gives the details of the statistical analysis implemented. Finally we conclude the paper by giving a conclusion in section 7.

II. Research Efforts Summary

As findings of a previously funded research project [7], two baseline HMM-based systems for phonemes and allophones [8, 9] were constructed using the mentioned speech database. The number of allophones in the speech database is 110 plus a silence unit which is counted as normal allophone indicating short pauses during the recitations, while the number of phonemes is 60, which represents almost half of the number of allophones. All speech units were modeled by an HMM with three emitting states for both levels to capture their acoustic properties. And for each state, a Gaussian Mixture Models (GMMs) were also associated to designate the characteristics of the sound portion at this state. The Mel-frequency cepstral coefficients (MFCCs) were used as cepstral acoustical features. For each Hamming window of 10 ms, a vector of 39 MFCCs was extracted. These coefficients are the first twelve MFCC plus their first and second derivatives to capture the sound's static features at this portion. Also, the energy plus its first and second derivatives were appended to identify the sound's dynamic features at the same portion. The hidden Markov model toolkit (HTK) was employed to train and test the HMMs for both systems. The word error rates (WERs) obtained for these recognizers were respectively 8% and 12% for phonemes and allophones.

Our current efforts focalized on the development of an elaborate system, by firstly considering the basic sounds and then looking for their distinctive features to determine which ones will be particularly helpful to well identify their phonological variation. To this end, we have adopted the speech database to be annotated in terms of basic phonemes. We mean by the basic phonemes the basic sounds without any phonological variation and even without considering the sounds gemination (the doubling). They are 32 phonemes. Their list and their associated codes are shown in the table 2.

The new version of the speech database was utilized in all efforts yet accomplished, including an HMM-based recognizer for basic Arabic sounds [10], an enhanced Arabic phonemes recognizer using duration modeling techniques [11] and an accurate HSMM-based system for Arabic phonemes recognition [12]. In the last implemented system for the basic Arabic phonemes [12], the average recognition rates obtained are about 99 %.

III. Background and Motivation

Automatic Speech recognition (ASR) seems to be actually in a blocked alley. Nearly all solutions are of the same general model [13]. The research focused on enhancing its performance by integrating supplementary elements. Such an approach yielded better results but it must be admitted that there is a limit which cannot be overrun without modification of the general scheme. The method based on hidden Markov models (HMMs) with features of fixed frames length has found its utility in numerous applications. However, it does not seem to be effective enough to transcribe properly any spoken language with a large vocabulary. There are several reasons. Some of them are very straightforward in their nature. The dictionary-based ASR system will never work correctly for out-of-dictionary words. Grammar models will not deal correctly with incorrectly spoken utterances while humans very often can.

ASR system tries to recognize speech via these matching techniques, while humans can easily understand it and adopt it to mistakes and unusual words. This causes the mentioned limit of the classical ASR approaches. The standard ASR approach is, indeed, based on guess and luck in few steps of its procedures. The inputted speech is segmented into frames without any motivated rules. HMM attempts to find the closest transcription on the basis of speech features which, indeed, a kind of guessing. Such approach works well enough for plainly spoken words with a limited vocabulary. Noise, the speaking rate and the large vocabulary cause many exclusions and data missing which HMM cannot deal with correctly. Another major problem is that people do not speak as carefully as they write, while we anticipate a transcription produced by an ASR system to be of the grade of our typed texts.

It has also to be admitted by both ordinary users and researchers, that when we speak we do not, at all times, follow grammar rules and, furthermore, the mistakes in pronunciation involve various exceptions independently of the dictionary size used. This is why adopting a hypothesis using related language rules and a limited dictionary does not always work satisfactorily. The same issues take place in the case of names, out-of-language words, and the mispronounced phonemes, etc. ASR system attempts to adopt the inputted speech to the language rules and the static vocabulary, which, in certain cases, leads to supplementary distortions and hence to degradation in system performance.

There is no straightforward solution for the above-described problems. In this work, we suggest the use of collected phoneme statistics in a target language in order to be used as, for instance, a support for the dictionary if there is a difficulty in associating matching features to one of the words to be recognized in the vocabulary.

The most outstanding research works carried out on continuous speech is based on statistical approaches specifically Hidden Markov Models (HMM). Many HMM-based ASR systems for continuous Arabic speech have reached various levels of recognition accuracy and encouraging performances which have been achieved [14-18]. The accuracy of recognition is usually measured by the correct percentage of recognized phonemes. The HMM-based ASR systems performance is affected by various factors including the existence of noise; the number of HMM states associated with each phoneme; the phoneme combination used and the phonemes length. Enhancing performance of the present ASR techniques needs the examination of these cited factors in order to localize and recognize the regions of enhancement.

Nonetheless, no fully statistical analysis at the phoneme level has been implemented on this speech database of classical Arabic sounds used in this work. Statistical analysis of Arabic phonemes gives a comprehensible vision of phonemes behavior and provides the capability to regulate this behavior by investigating the gathered statistics. For example, the frequency of a specific phoneme in a speech database can be employed to correct its misrecognition during the decoding process. This means replacing this misrecognized phoneme by the highest probably one.

Furthermore, the average duration of a particular phoneme can also be utilized to estimate the number of HMM states that are most appropriate for recognizing it. Additional statistical information such as mode (the midst value in a set of values) and median (the most frequent value in a set of values) are advantageous in addressing the misrecognized phonemes during the decoding process. In this paper, we present a full statistical analysis of Arabic phonemes which can be employed for the purpose of enhancing performance of our baseline HMM-based systems by reducing the word error rate (WER) factor.

IV. Speech Database of Sounds

The Arabic language is the official language of about 300 million speakers around the world. It is the religious language of all Muslims around the world, regardless of their native language. It is the official language in all Arab countries and the 6th most widely utilized language in terms of first language speakers. Arabic can be categorized into two main variants: Classical Arabic (CA) and Modern Standard Arabic (MSA). CA is an old literary form of Arabic, which is the most formal type and is the language of the Holy Quran and the old Arabic poetry. MSA is the current standard form of Arabic, which is utilized in official communications in Arabic countries, broadcast news, formal speeches, etc. Although there is no big difference between today's Arabic (MSA) and that spoken by the early Arabs (CA), due to the fact that Arabic is one of the most stable languages throughout history, yet there are some idiosyncrasies as to the way of pronunciation.

One of the main barriers faced by the development of ASR applications for Arabic speech is the rarity of suitable sound databases commonly required for training and testing statistical models. This problem is seriously approached when dealing with classical Arabic language since most of the corpora available nowadays are specifically oriented towards what is known as Modern Standard Arabic (MSA) and its sub-forms (i.e. dialects). To remedy this problem and to assist the development of ASR applications for classical Arabic language, a speech database covering all classical Arabic sounds was designed on the basis of Quranic recitations. The speech corpus was developed in a previously funded project by Al-Imam Muhammad ibn Saud Islamic University in Saudi Arabia with the support of King Abed Al-Aziz City for Science

and Technology (KACST). Because of the difficulty of developing this kind of corpora, only a part of the Holy Quran was regarded. Recitations of ten male speakers were recorded in an appropriate environment under the supervision of an expert of the holy Quran pronunciation rules (called Tajweed); more than eight hours of speech were achieved [19-21]. Each audio file is a Quranic verse or a portion of it for long verses where the speaker must take a long breath.

In order to have a speech database useful for many goals, speech signals were manually and accurately segmented into three levels: word, phoneme and allophone. A new labeling system was proposed to annotate the speech segments [16] because the labeling systems available (e.g. IPA, SAMPA, BEEP, etc.) were not able to cover all Arabic sounds. However, the speech database consists of 44.1 KHz wav files of 16 millisecond utterances over its corresponding MFCC feature files, label files and TextGrids files.

Table I lists for each speaker, the number of sound files, their size and duration. The list of basic Arabic phonemes and their associated codes are shown in table II.

TABLE I. SOUND FILES AND THEIR DURATION BY SPEAKERS

Speaker Number	Speaker Initials	Number of Sound Files	Duration (minutes)	Size (MB)
1	AAH	600	49.36	249
2	AAS	590	52.09	261
3	AMS	612	45.78	229
4	ANS	597	49.72	250
5	BAN	585	54.75	276
6	FFA	578	44.11	220
7	HSS	601	49.76	251
8	MAS	580	46.24	232
9	MAZ	608	51.47	258
10	SKG	584	44.29	220
Total		5935	487.53 (8h, 8m)	2446

TABLE II. LIST OF BASIC ARABIC PHONEMES AND THEIR CODES

Arabic Orthography		Label	Arabic Orthography		Label
ـَ	فتحة	as10	ص	صاد	sb10
ـُ	ضمة	us10	ض	ضاد	db10
ـِ	كسرة	is10	ط	طاء	tb10
ء	همزة	hz10	ظ	ظاء	zb10
ب	باء	bs10	ع	عين	cs10
ت	تاء	ts10	غ	غين	gs10
ث	ثاء	vs10	ف	فاء	fs10
ج	جيم	jb10	ق	قاف	qs10
ح	حاء	hb10	ك	كاف	ks10
خ	خاء	xs10	ل	لام	ls10
د	دال	ds10	م	ميم	ms10
ذ	ذال	vb10	ن	نون	ns10
ر	راء	rs10	هـ	هاء	hs10
ز	زاء	zs10	و	واو	ws10
س	سين	ss10	ي	ياء	ys10
ش	شين	js10	صامت		sil

In addition, the speech database contains a list of 60 Arabic phonemes, an Arabic dictionary, a list of all unrepeated words included in the whole eight hours speech database and other useful files needed for the recognizer development.

V. STATISTICS EXTRACTION METHOD

To extract statistics from the speech database, a computer program was designed using MATLAB programming language developed by MathWorks [22]. The occurrence probability of each basic phoneme, frequency of occurrence of basic phoneme, mean duration, Min and Max durations for each basic phoneme, mode and the median of duration for each basic phoneme were calculated. Durations are computed on the basis of phonemes boundary extracted from TextGrids files attached withal the speech database Sound.

These gathered statistics are displayed in Table 3 (see Table III) which also shows the labels used for every basic phoneme in the speech database. Fig. 1 shows the mean of basic phonemes durations measured in second. The frequency of each basic phoneme in the whole database is shown in Fig. 2. For an in-depth analysis of the collected statistic and for the purpose to have extra information about the characteristics of the basic Arabic phonemes, useful graphs are depicted in Figures 3, 4,5 and 6.

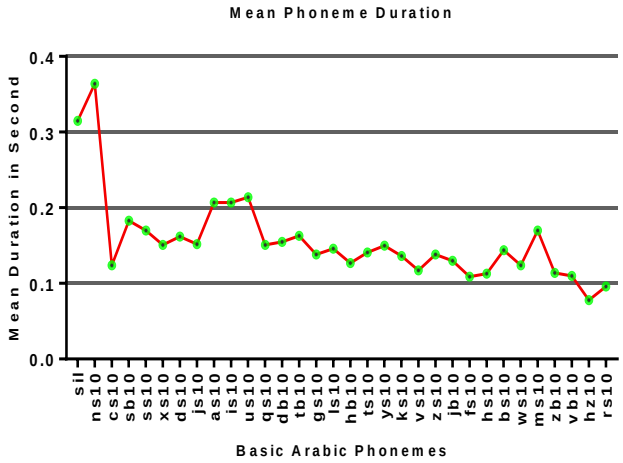

Fig. 1. Mean Duration of the Basic Arabic Phonemes

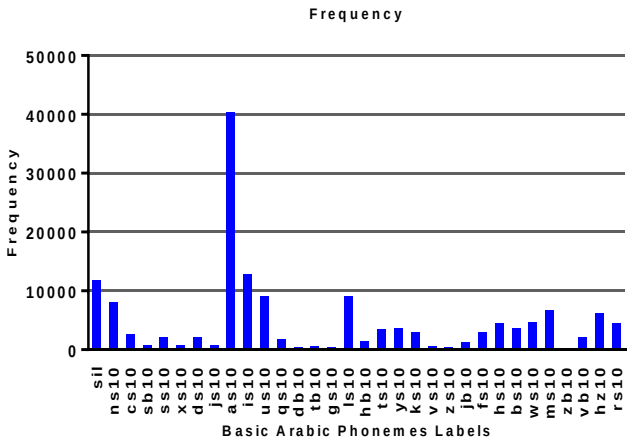

Fig. 2. Basic Arabic Phonemes Frequencies

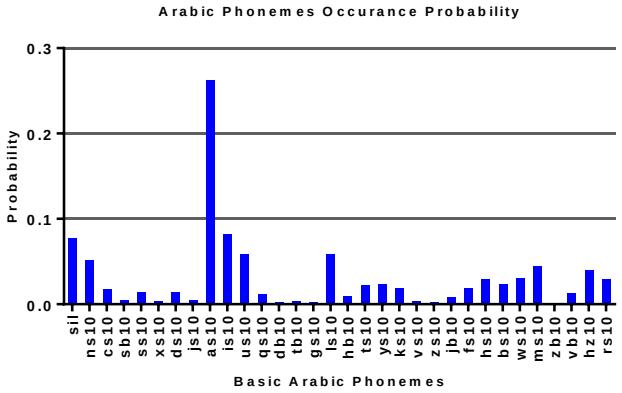

Fig. 3. Basic Arabic Phonemes Occurrence Probability

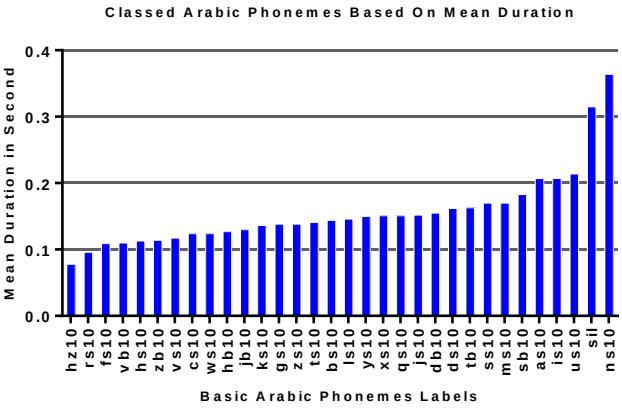

Fig. 4. Sorted Basic Arabic Phonemes based on their Means

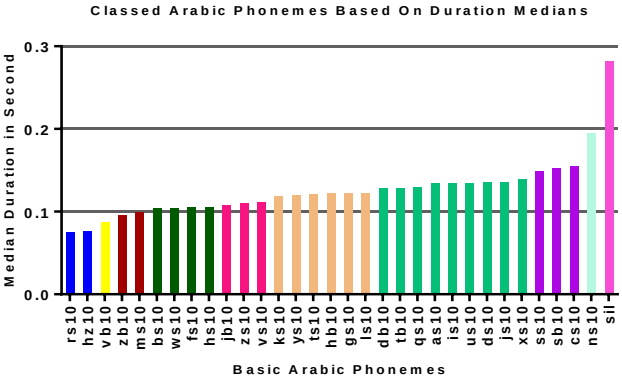

Fig. 5. Sorted Basic Arabic Phonemes based on their Medians

Fig. 3 shows the occurrence probability of the basic Arabic phonemes in the whole speech database. This useful graph will serve in defining the probability of missing phonemes during the decoding process. However, we noted that the phoneme "sil" denoting the silence regardless of its occurring places in the speech database is included in all depicted graphs.

In interesting outcome which is apparent from Fig. 4 proves that basic phonemes having equal or approximate mean values can be grouped into clusters. we assume that these clusters will be helpful for the purpose of enhancing performance of the baseline recognizer as we will evoke in the next sections. Basic phoneme duration medians give a clearly view of those clusters. Classes of the phonemes groups are being differentiated from each other and a clear parting among phoneme groups becomes more obvious, as seen in Fig. 5.

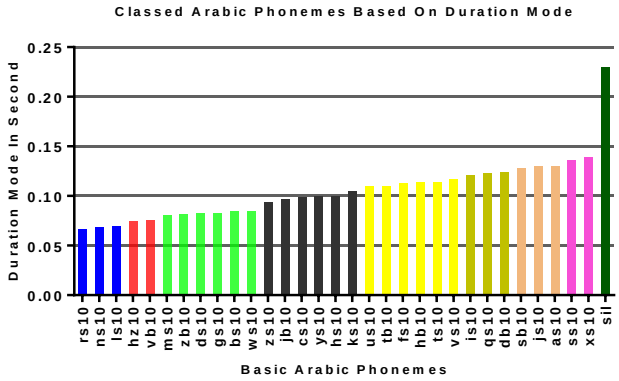

Fig. 6. Sorted Basic Arabic Phonemes based on their Modes

Another significant graph is the one demonstrating the most frequent duration value of all occurrences of a basic phoneme appearing in the "CA Sound Database". This is referred as the mode, and is displayed in Fig. 6.

VI. Statistics Analysis

When taking a look at the previous tables and graphs, we find that each basic phoneme occurs with various frequencies, the highest frequent ones are "as10" (فتحة), is10" (كسرة) and "us10" (ضمة), respectively, which designate the Arabic vowels. Otherwise the smallest frequent ones are "zb10" (حرف الظاء), "gs10" (حرف الغين), and "zs10" (حرف الزاء), respectively, ignoring the phoneme denoting the silence "sil" (صامت). From the results shown in Figures 2 and 3; it seems clear that when a phoneme is missed throughout the decoding process, phoneme "as10" is automatically the most probable one replacing it. Generally, the results concluded from Fig. 3 can be employed to correct the pronunciations for a misrecognized phoneme in spoken utterances during the recognition phase. The use of this information seems useful in enhancing the baseline system performance.

Fig. 4 illustrates the entire basic Arabic phonemes sorted on the basis of their average durations. From this Figure, we can clearly show the behavior of the basic phoneme durations through the whole speech database. Thus, the figure provides an explicit idea about the average duration of each phoneme, which means that a basic phoneme clusters being distinguished from it. For example, the basic phonemes "hz10" and "rs10" form the first cluster. The second cluster includes: "vb10","fs10" and "hs10". The vowels form the last cluster in terms of the highest average durations. Usually, knowing the average length of a specific phoneme in a speech database can be utilized for estimating the appropriate number of the HMM states that represent it, which generally accelerate the estimation period and hence enhance the accuracy of recognition.

In Fig. 5 and Fig. 6, median and mode durations for each basic phoneme are displayed, where the basic phonemes clusters appear clearly. The outcomes of both figures could be helpful to make the correct decision in dealing with either misrecognized or missed phonemes. It means that replacing them with the near median or mode phoneme.

VII. CONCLUSION

In this paper, we have presented a collection of statistical data for Basic Arabic phonemes helpful in enhancing HMM-based automatic speech recognition systems performance. In the literature, the duration of phonemes is regarded as major distinctive feature characterizing the voice of a speaker. Knowing the duration of a particular phoneme in a spoken utterances can be utilized to estimate the length of the HMM chain describing it, which in consequence improves the system performance. These investigations were performed using a particular speech database of Quranic sounds including more than eight hours of speech and ten different male speakers. The numerical values are extracted using a computer program designed for this purpose. A discussion of these results with interpretations was also presented and reported graphically. Dividing phonemes into clusters on the basis of their median of the durations can help in decreasing the search for the appropriate phoneme during the decoding process, which in consequence increases system performance. Collected statistics provided can also be used to build or propose other techniques for phonemes classifications. While the probability distributions in HMM-based ASR systems are usually estimated with the Expectation-Maximization iterative algorithm, the statistics provided can be utilized as an initial condition for the estimation procedure, and, thus, speed up its execution time, or can also be utilized as a wanted model itself. We believe that the absence of necessary numerical data denoting, particularly, the basic Arabic phonemes behavior in classical Arabic language like those reported here gives an added value to the presented work. However, our future steps will focus on incorporating these statistics explicitly into HMMs in order to overcoming the classical HMM's weakness and, hence, improve HMM-based systems performance.

ACKNOWLEDGMENT

The presented work utilizes the results (Classical Arabic Sound Database) of a project previously funded by King Abed Al-Aziz City for Science and Technology (KACST) in Saudi Arabia under grant number “AT – 25 – 113”.

REFERENCES

[1] D. Jurafsky and J. H. Martin, Speech and Language Processing, 2nd ed., Pearson Prentice Hall, 2009.

[2] G. Zweig and P. Nguyen, “A segmental CRF approach to large vocabulary continuous speech eecognition,” Proc.of IEEE ASRU, 2009.

[3] H. Sakoe, Two-level DP-matching - a dynamic programming-based pattern matching algorithm for connected word recognition, Readings in Speech Recognition, Morgan Kaufmann Publishers Inc, pp. 180-186, 1990.

[4] H. Jiang, “Discriminative training for automatic speech recognition: A survey,” Computer Speech & Language, Comput. Speech, vol. 24, no. 4, pp. 589–608, 2010.

[5] L. Deng and X. Li, "Machine Learning paradigms for speech recognition: An overview,” IEEE Trans. Audio, Speech, Lang. Process., vol. 21, no. 5, pp. 1060–1089, May 2013.

[6] I, Oparin, Language Models for Automatic Speech Recognition Of inflectional Languages. PhD Thesis, University of West Bohemia, Plzen, Czech Republic (2009).

[7] Y.O.M. Elhadj, I.A. Alsughayeir, M. Alghamdi, M. Alkanhal, Y.M. Ohali, A.M. Alansari, Computerized teaching of the Holy Quran (in Arabic), Final Technical Report, King Abdulaziz City for Sciences and Technology (KACST), Riyadh, KSA,2012.

[8] Y.O.M. Elhadj, M. Alghamdi, and M. Alkanhal, “Phoneme-Based Recognizer to Assist Reading the Holy Quran,”, Recent Advances in Intelligent Informatics, Advances in Intelligent Systems and Computing, Springer, pp.141-152,2014.

[9] Y.O.M. Elhadj, M. Alghamdi, and M. Alkanhal, “Approach for Recognizing Allophonic Sounds of the Classical Arabic Based on Quran Recitations,”, Theory and Practice of Natural Computing, Lecture Notes in Computer Science, Springer, pp. 57-67, 2013.

[10] Y.O.M. Elhadj, Mohamed .O.M. Khelifa, A. Yousfi and M. Belkasmi. “An Accurate Recognizer for Basic Arabic Sounds,” ARPN Journal of Engineering and Applied Sciences, vol. 11, no. 5, pp. 3239- 3243, Mar. 2016.

[11] Mohamed O.M. Khelifa, Y.O.M. Elhadj, Y. Abdellah and M. Belkasmi, “Enhancing Arabic Phoneme Recognizer using Duration Modeling Techniques,”, in proc. of Fourth International Conference on Advances in Computing, Electronics and Communication - ACEC 2016, Dec 15, 2016, Rome-Italy.

[12] Mohamed O.M. Khelifa, Y.O.M. Elhadj, Y. Abdellah and M. Belkasmi, “An Accurate HSMM-based System for Arabic phonemes Recognition,” in proc. of The IEEE Ninth International conference on Advanced Computational Intelligence (ICACI 2017), Feb. 2, 2017, Doha, Qatar.

[13] S. Young, Large Vocabulary Continuous Speech Recognition: a Review, IEEE Signal Processing Magazine 13(5), pp. 45-57, 1996.

[14] Ali, A. et al., “A Complete KALDI Recipe for Building Arabic Speech Recognition Systems”, Spoken Language Technology Workshop (SLT), IEEE, 2014.

[15] Khalid, A. et al., "Arabic Phonemes Transcription using Data Driven,"The International Arab Journal of Information Technology, Vol. 12, No. 3, May 2015.

[16] Speaker-dependant continuous Arabic speech recognition. M.Sc. thesis, King Saud University, 2001.

[17] Hyassat H, Abu Zitar, “Arabic speech recognition using SPHINX engine,”, Int J Speech Tech 9(3–4):133–150, 2008.

[18] Azmi, M. et al., “Syllable-based automatic Arabic speech recognition in noisy-telephone channel,”, In: WSEAS transactions on signal processing proceedings, World Scientific and Engineering Academy and Society (WSEAS), vol 4, issue 4, pp 211–220, 2008.

[19] Y.O.M. Elhadj, M. et al., Design and Development of a High Quality Speech Corpus for Classical Arabic. Submitted for publication to the Language Resources and Evalauation Journal (LREV).

[20] Y.O.M. Elhadj, M. et al., Sound Corpus of a part of the noble Quran (in Arabic). Proc. of the International Conference on the Glorious Quran and Contemporary Technologies, King Fahd Complex for the Printing of the Holy Quran, Almadinah, Saudi Arabia, October 13-15, 2009.

[21] Y.O.M. Elhadj. Preparation of speech database with perfect reading of the last part of the Holly Quran (in Arabic). Proc. of the 3rd IEEE International Conference on Arabic Language Processing (CITAL'09), pp: 5-8, Rabat, Morocco, May 4-5, 2009.

[22] MATLAB and Statistics Toolbox Release 2013a The MathWorks, Inc., Natick, Massachusetts, United States.

TABLE III. THE BASIC ARABIC PHONEMES STATISTICS

Basic Arabic Phonemes	Labels	Frequency of occurrence	Min duration in second	Max duration in second	Mean-duration in second	Mode	Median	Probability of occurrence
صامت	sil	11875	0.022	8.576	0.315	0.230	0.282	0.077
نون	ns10	8160	0.021	1.458	0.364	0.068	0.195	0.052
عين	cs10	2700	0.033	0.420	0.124	0.099	0.155	0.017
صاد	sb10	838	0.079	0.388	0.183	0.128	0.153	0.005
سين	ss10	2175	0.071	0.384	0.170	0.136	0.149	0.014
خاء	xs10	770	0.072	0.420	0.151	0.139	0.139	0.004
دال	ds10	2190	0.039	0.433	0.162	0.083	0.136	0.014
شين	js10	867	0.080	0.478	0.152	0.130	0.136	0.005
فتحة	as10	40396	0.011	3.343	0.207	0.130	0.135	0.262
كسرة	is10	12755	0.030	1.833	0.207	0.121	0.135	0.082
ضمة	us10	9110	0.029	1.739	0.214	0.110	0.135	0.059
قاف	qs10	1870	0.080	0.792	0.151	0.123	0.130	0.012
ضاد	db10	443	0.021	0.629	0.155	0.124	0.128	0.002
طاء	tb10	560	0.073	0.464	0.163	0.110	0.128	0.003
غين	gs10	410	0.049	0.387	0.138	0.083	0.123	0.002
لام	ls10	9066	0.015	0.767	0.146	0.069	0.123	0.058
حاء	hb10	1457	0.050	0.335	0.127	0.114	0.122	0.009
تاء	ts10	3483	0.019	0.959	0.141	0.114	0.121	0.022
ياء	ys10	3677	0.019	1.392	0.150	0.100	0.120	0.023
كاف	ks10	3040	0.028	0.480	0.136	0.105	0.119	0.019
ثاء	vs10	600	0.032	0.311	0.117	0.117	0.112	0.003
زاء	zs10	440	0.060	0.352	0.138	0.094	0.111	0.002
جيم	jb10	1240	0.015	0.428	0.130	0.097	0.108	0.008
فاء	fs10	3020	0.016	0.369	0.109	0.113	0.105	0.019
هاء	hs10	4559	0.029	0.376	0.113	0.100	0.105	0.029
باء	bs10	3739	0.012	0.654	0.144	0.085	0.104	0.024
واو	ws10	4647	0.016	1.021	0.124	0.085	0.104	0.030
ميم	ms10	6825	0.027	1.640	0.170	0.080	0.099	0.044
ظاء	zb10	176	0.054	0.360	0.114	0.082	0.096	0.001
ذال	vb10	2091	0.031	0.371	0.110	0.076	0.087	0.013
همزة	hz10	6281	0.008	0.295	0.078	0.074	0.076	0.040
راء	rs10	4620	0.014	0.403	0.096	0.066	0.075	0.029