Journal of Theoretical and Applied Information Technology
31st July 2019. Vol.97. No 14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ARABIC SPEECH RECOGNITION SYSTEM AND PHONETIC DICTIONARY

[1] IBRAHIM EL-HENAWY, [2]MARWA ABO-ELAZM

[1, 2] Computer Science Department, Faculty of Computer Science and Information Technology, Zagazig University, Zagazig Egypt.

E-mail: [1]henawy2000@yahoo.com, [2]marwa_abdella@yahoo.com

ABSTRACT

Automatic Speech Recognition (ASR) is a technology by which a computer can distinguish the words that a person talks into a microphone or a telephone. It is a key technology for many applications. It is gaining a growing role in a variety of application such as telephone communication with information systems, automatic query answering, hands free operation control as in airplanes and car, speech-to-text transcription, etc.

ASR can help also, handicapped people to interact with society. Due to the ASR importance many systems are developed, the most common are: Dragon Naturally Speaking, IBM via voice, Microsoft SAPI. Also, many speech recognizing open source systems are available too.

One of the main obstacles challenge the development of ASR applications for Arabic speech is the rarity of suitable sound databases commonly required for training and testing statistical models.

The pronunciation (or phonetic) dictionary is an important ASR component that serves as an intermediate between language models and acoustic models in ASR systems. It holds a subset of the words available in the language and the pronunciation variants of each word in terms of sequences of the phonemes existing in the acoustic models. The development of perfect Automatic Arabic Speech Recognition (ASR) systems is challenged with the accuracy of the phonetic dictionary.

Keywords: *Speech Recognition Systems, Arabic Language, Phonetic Dictionary, Pattern Recognition*

1. INTRODUCTION

This Speech recognition can be defined as the process of transforming the audio signal, taken by a microphone or a telephone, to a set of words. The using of speech recognition software as a consumer tool start to grow in the 1990s.

Arabic as one of the six formal languages of the United Nations (UN) is one of the most broadly spoken languages in the world. Statistics indicate that it is the first language (mother-tongue) of 206 million native speakers classified as the fourth after Mandarin, Spanish and English [1]. Despite the importance of Arabic language research effort on Arabic Automatic Speech Recognition (ASR) is unfortunately still insufficient. Many issues for Arabic language that need to be addressed to catch up with the progress of other language

The development of perfect Automatic Arabic Speech Recognition (ASR) systems is challenged with two major issues. One of them is dicritization where diacritic marks refer to vowel phonemes in the nominated words. Arabic texts are not fully diacritized: that make the short strokes placed above or below the consonant, that represent the vowel following this consonant, are usually absent. This bounds the availability of Arabic ASR training material. The lack of this information leads to huge number of similar word forms, and consequently, the language model is less predictable. The other problem is associated with the morphological complexity since Arabic has a large potential of word forms that increases the out-vocabulary rate.

2. ARABIC SPEECH RECOGNITION

There are 3types of Arabic language[2] that is Classical Arabic(CA), the Qur'an language , and it

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

is used mainly for reciting and reading Islamic holy text, Modern Standard Arabic (MSA) , used in TV and the news the "common language" used by speakers of different dialects, and Spoken Arabic (dialect), the dialects of Arabic can be separated into two collections: Western Arabic, which includes the dialects spoken in Libya, Morocco, Algeria and Tunisia, and Eastern Arabic, which can be also subdivided into Gulf Arabic, Egyptian, and Levantine. These various dialects differ significantly from each other and from Modern Standard Arabic.

Some of relationships exist between MSA and CA, in which MSA depend on the CA syntactically, morphologically, and phonologically but lexically MSA is a modernized version of CA[3] . Despite the importance of automatic speech recognition and its use in many applications, its research isn't adequate in many sides and research effort on Arabic ASR still requires more weight worldwide [4].

3. CHARACTERISTICS OF ARABIC LANGUAGE

The Arabic alphabetic consists of 28 letters. Various extra letters are used in Arabic when we write place names or foreign words containing sounds which do not exist in Standard Arabic, such as /v/ or /g/. Arabic words are written from right to left in horizontal lines, numerals are written from left to right. Most of letters change its shape according to whether they written at the beginning end or middle of a word, or on their own. Numerous other pronunciation phenomena are indicated by diacritics, such as word-final adverbial markers that add /n/ to the pronunciation, that indicated by "tanween" , and consonant doubling , which is specified by the "shadda" sign. Furthermore, beginners' texts, such as schoolbooks of children, is discretized, they begin to be removed as soon as the learner has the language knowledge. If the diacritics not exist, it may lead to considerable lexical ambiguity. These ambiguities must be resolved by contextual information, which hypothesizes knowledge of the language. Without this knowledge, it is difficult to determine how to pronounce a non-diacritized text [4].

4. ARABIC SPEECH RECOGNITION SYSTEM DIFFICULTIES

The recognition system of Arabic language face several difficulties [5] which are:

- **Word Knowledge***: the* meaning of a word is required that make the intended speech recognized correctly.
- **Patterns Variability Caused By Dialectical Differences:** Arab countries have a variety of dialects and also different regions in one country have dialectical difference between them making the word to be pronounced in many ways. This variability in the pronunciation makes the recognition of a word more difficult.
- **Co articulation effects:** The acoustic recognition of a phoneme depends on the acoustic context in which it exist. This is typically called Co articulation. Thus, the neighboring phonemes affect the acoustic feature of a phoneme. The position of a phoneme in a word and the position of this word in a sentence also have an effect. These acoustic features are very different compared to isolated phonemes, since the articulatory organs move much in continuous speech than in isolated utterances.
- **Diacritization***:* Diacritics play an essential rule in textual Arabic material. Most of Arabic texts not include diacritics making the pronunciation of words more ambiguous.

There are also several challenges facing ASR system such as stuttering, coughing, false starts, dis-fluency, pitch, and repetitions but have moreover the challenges that arise from the pace of speech as described in [6].

5. SPEECH RECOGNITION SYSTEMS CLASSIFICATION

Speech recognition systems are classified according to the number of speaker into two types, speaker-dependent and speaker-independent recognition systems. Speaker-dependent systems are designed using one speaker. They generally are more accurate for this speaker, but mush less accurate for others, whoever Speaker-independent systems are designed for a variety of speakers. In speaker–independent the system recognizes speech of many individuals without training but in speaker-dependent it has to be trained for a particular voice. Speaker adaptive system starts as speaker-independent systems and using training techniques to adapt speaker voice and increasing their recognition accuracy [7].

Speech recognition systems are classified based on the recognized utterance into isolated word recognition system which recognizes single utterance, connected words system where separate utterances to be "run-together" with a little silence between them can be recognized, Continuous speech recognition system in which users can communicate almost naturally, Spontaneous speech recognition system recognizes the natural speech.

6. AUTOMATIC SPEECH RECOGNITION SYSTEM PHASES

Speech recognition consists of two main phases front-end (feature extraction) and back-end (pattern recognition) phase.

There are different types of features that represent the acoustic data and the selection of the best parametric representation of acoustic data is an important task in the development of the speech recognition system. There are many speech features that dominated the speech and speaker recognition areas in consequent periods: Perceptual Linear-Predictive coefficients (PLP), Reflection coefficients(RC) , Linear prediction coefficients (LPC), Linear prediction cepstrum coefficients (LPCC), Linear frequency cepstrum coefficients(LFCC), and Mel frequency cepstrum coefficients(MFCC) which are Fourier based features. MFCC provide better parameters and are commonly used as a feature extraction method in recent research [8][9].

6.1 Mel frequency cepstrum coefficients(MFCC)

MFCC algorithm take the speech signal preform certin processing and produce a set of vector that uniquedescribe each sample. Figure 1 illustrates the steps of feature extraction mechanism.

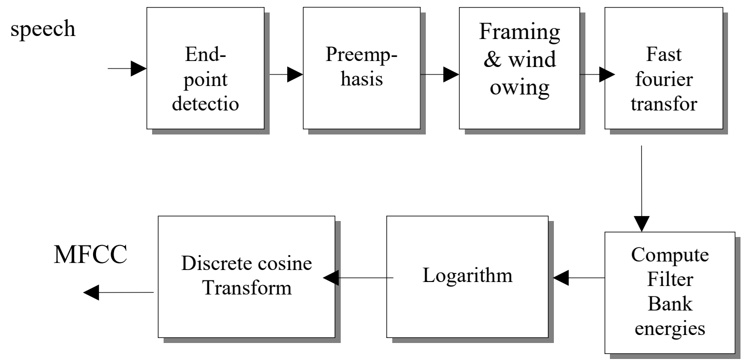

Figure 1 Steps Of Feature Extraction Mechanism

Endpoint detection is the vital step for Speech and Speaker Recognition applications. Probability Density Function (PDF) of the background noise and a Linear Pattern Classifier for classification of Voiced part of a speech from silence or unvoiced part are applied to the signal. The previous work shows enhanced silence removal and end point detection than conventional Zero Crossing Rate (ZCR) and Short Time Energy (STE) function methods[10].

The digitized signal is passed through a finite impulse response filter called Pre_emphasis filter ($H(z) = 1- 0.97z$ -1) that is used to flatten the spectra tilt and reduce the effects of the glottal pulses and radiation impedance and to focus on the spectral properties of the vocal tract[11]

The signal is divided into frames to minimize the discontinuity and preventing spectral leakage of a signal at the beginning and end of each frame, and the assumption that a signal within a frame is stationary.

After the frame blocking is done a hamming window is applied to each frame using Eq [1]. This

Journal of Theoretical and Applied Information Technology
31st July 2019. Vol.97. No 14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

window is to reduce the signal discontinuity at the ends of each block [12].

$$Wh(n) = 0.54 - 0.46\cos(2\pi n)/(N-1)) \quad (1)$$

For each frame a fast fourier transform is executed to obtain the magnitude frequency for each frame.
The discrete cosine transform (DCT) is performed on the output of the logarithm filter-bank to generate cepstral parameters using Eq [2].
DCT produce feature that are less correlated than log energies and permits a reduction in the size of feature vector while enabling a higher frequency resolution by appropriate choice of the number of filters [13] .

$$Cj = \sum_{i=1}^{M} X_i \cdot \cos\left(j \cdot (i-0.5) \cdot \frac{\pi}{M}\right), \; j = 0,1,.,J-1 \quad (2)$$

M is the number of filters in the filter-bank and J is the number of MFCC coefficients that are needed. For speech recognition J = 13 is a widely accepted value.

6.2 The back end phase

There are 3 important model in the back-end phase that are Acoustic model, language model and phonetic dictionary. The development of the Acoustic model is for distinguishing the spoken phoneme. The creation of the model require the using of the recorded audios of the speech a long with their text scripts and next compiling them into a statistical representation of sounds which construct the words.

The Language modeling designed for identifying the relations between the sentence words with the aid of pronunciation dictionary. It is the grammar of the language [14]. ASR systems make use of n-gram language models to guide the search for correct word sequence by predicting the likelihood of the nth word on the basis of the $n-1$ antecedent's words. For a set of words W=w1, w2, w3, …, w_n the language model computes there probability using equation 3.

$$P(W) = P(w_1, w_2, w_3, \ldots, w_n) = \prod_{i=1}^{n} p(w_i) | P(w_1, \ldots P(w_{i-1}) \quad (3)$$

phonetic dictionary works as an intermediary between the Acoustic Model and the Language Model in speech recognition systems. The creation of a good dictionary has a large effect on the accuracy of the automatic speech recognition systems [15].

7. PHONETIC DICTIONARY GENERATION

Phonetic dictionary consist of the orthographical symbols of words and their phonetic or phonemic pronunciation variants. It contains a subset of the words that can be exit in the language and its pronunciation phonemes or allophones the acoustic model can have. It also defines the set of valid phoneme sequences and therefor is a key element in guiding the search space of a recognizer. The quality with which it maps the orthography of a word to its pronunciation according to every speaker voice has a great influence on the performance of the recognition system in two ways [16]. First In the training phase if the map of a ward to its phonetic unit is incorrect, it will contaminate the acoustic models. The models actual acoustic representation will not be as accurately as if they were trained using the correct data. Second With the correct training of the acoustic model the false mapping make hypothesis scoring faked because of applying the incorrect model to the calculation.

The creation of a phonetic dictionary is a complex task. The creation can be done manually by the language experts for the language that has a large number of pronunciation exceptions such as English but this approach can be very expensive and take a large time especially for large vocabulary recognizers. Also the creation can be automatically created for Arabic language because the pronunciation of Arabic text follows specific rules when the text is fully diacritized. Many of these pronunciation rules can be found in [17][18].

In Arabic Automatic Speech Recognition Systems, Romanized phonemes or phonemes composed of Arabic letters can be used to represent the

Journal of Theoretical and Applied Information Technology
31st July 2019. Vol.97. No 14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

pronounced letters and words in training the system. Table 1 show sample from arabic phonem set

Table 1 Arabic Letter Phoneme Set

number	letter	Description
1	بَ	The letter BA with diacritical mark FATHA
2	بِ	The letter BA with diacritical mark KASSRA
3	بُ	The letter BA with diacritical mark DHAMMA
4	بْ	The letter BA with diacritical mark SOKON
5	بَّ	The letter BA with diacritical mark SHADDA AND FATHA
6	بِّ	The letter BA with diacritical mark SHADDA AND KASSRA
7	بُّ	The letter BA with diacritical mark SHADDA AND DHAMMA

Although most of AASRS use the Romanized phonemes shown in table 2 for Arabic because it obtained good results , it isn't natural, reading it is difficult, and consume a lot of time to prepare and generate the phonemes used in the audio files transcription for the training of the language [19].

Table 2 Romanized Phoneme Set

Number	phoneme	Romanized phoneme	Description
1	ب	/B/	Arabic Voiced Consonant BA
2	ت	/t/	Arabic Voiced Consonant TA
3	ث	/th/	Arabic Consonant THA
4	ج	/g/	Arabic Consonant GEEM
5	ح	/h/	Arabic Consonant HA
6	خ	/kh/	Arabic Consonant KHA

There are two approaches in constructing pronunciation dictionary for Arabic language, the first deal with vowels as independent phonemes from consonant phonemes so there are phoneme for short vowels (FATHA, KASSRA, DHAMMA)and phoneme for long vowels. The others say no phonemes for vowels, it is part of consonant phonemes because vowels can't be pronounced alone, and they must be preceded by the consonant letters so they shouldn't be represented by independent phoneme [20]. Table 3 show the phoneme set used in the creation of the phonetic dictionary of the first approach. They choose phonemes similar to the phonemes used in English Automatic Speech Recognition that produced better recognition rate.

Table 3 First Approach Phoneme Set

Number	Arabic phoneme	Romanized phoneme set	Description
1	ـَ	/AE/	diacritical marks FATHA
2	اَ	/AE:/	long vowel of AE
3	ـَ	/AA/	the pharyngealizd allophone of /AE/
4	ـَ	/AA:/	Long version of AA
5	ـَ	/AH/	Emphatic Version of /AE/
6	ـَ	/AH:/	Long version of AH
7	ـٌ	/UH/	diacritical marks DHAMMA
8	وُ	/UW/	Long vowel UH
9	ـُ	/UX/	the pharyngealized allophone of /UH/
10	*ِ	/IH/	diacritical marks KASSRA
11	يِ	/IY/	Long vowel of IY
12	*ِ	/IX/	the pharyngealizd allophone of /IH/
13	*ِ	/IX:/	Long version

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

			of IX
14	وَ	/AW/	A Diphthong of both /AE/ and /UH/
15	يَ	/AY/	A Diphthong of both /AE/ and /IH/
16	ء	/E/	Hamza
17	ب	/B/	Arabic consonant BA
18	ت	/T/	Arabic consonant TA
19	ث	/TH/	Arabic consonant THA
20	ج	/JH/	Arabic consonant GEEM
21	ح	/HH/	Arabic consonant HA
22	خ	/KH/	Arabic consonant KHA
23	د	/D/	Arabic consonant DAL
24	ذ	/DH/	Arabic consonant ZAL
25	ر	/R/	Arabic consonant RA
26	ز	/Z/	Arabic consonant ZA
27	س	/S/	Arabic consonant SEEN
28	ش	/SH/	Arabic consonant SHEEN
29	ص	/SS/	Arabic consonant SAD
30	ض	/DD/	Arabic consonant DAD
31	ط	TT	Arabic consonant TA
32	ظ	DH2	Arabic consonant THA
33	ع	AI	Arabic consonant AIN
34	غ	GH	Arabic consonant GHAIN
35	ف	F	Arabic consonant FA
36	ق	Q	Arabic consonant QAF
37	ك	K	Arabic consonant KAF
38	ل	L	Arabic consonant LAM
39	م	M	Arabic consonant MEM
40	ن	N	Arabic consonant NON
41	ه	H	Arabic consonant HA
42	و	W	Arabic Semi-vowel WAW
43	ى	Y	Arabic Semi-vowel YA

Another phonetic dictionary is created and achieved better recognition rates that depend on: the vowels have no independent phoneme only consonants have and vowels are part of them [20]. There are 28 Arabic letter each letter has 7 pronunciation so there are 196 phoneme as shown in table 4.

Table 4 Second Approach Phoneme Set

Number	Arabic phoneme	Romanized phoneme set	Description
1	ب	/B/	The letter BA without diacritical mark

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2	بَ	/B-/	The letter BA with diacritical mark FATHA
3	بَا	/B--/	The letter BA with diacritical mark FATHA with ALIF
4	بِ	/B+/	The letter BA with diacritical mark KASRA
5	بِى	/B++/	The letter BA with diacritical mark KASRA with YA
6	بُ	/B*/	The letter BA with diacritical mark DAMMA
7	بُو	/B**/	The letter BA with diacritical mark DAMMA with WAW
. . .	. . .	. . .	. . .
183	و	/w/	The letter WAW without diacritical mark
184	وَ	/w-/	The letter WAW with diacritical mark FATHA
185	وَا	/w--/	The letter WAW with diacritical mark FATHA with ALIF
186	وِ	/w+/	The letter WAW with diacritical mark KASRA
187	وِى	/w++/	The letter WAW with diacritical mark KASRA with YA
188	وُ	/w*/	The letter WAW with diacritical mark DAMA
189	وُو	/w**/	The letter WAW with diacritical mark DAMMA with waw
190	ي	/y/	The letter YA without diacritical mark
191	يَ	/y-/	The letter YA with diacritical mark FATHA
192	يَا	/y--/	The letter YA with diacritical mark FATHA WITH ALEF
193	يِ	/y+/	The letter YA with diacritical mark KASRA
194	يِى	/y++/	The letter YA with diacritical mark KASRA with YA
195	يُ	/y*/	The letter YA with diacritical mark DAMMA
196	يُو	/y**/	The letter YA with diacritical mark DAMMA with WAW

Using the previous phoneme set and a set of rules that are developed in [21], a phonetic pronunciation for each Arabic word is generated. Sample from generated dictionary is indicated below

آ باَر E AE: B AE: R IX N
(2) آ باَر E AE: B AA: R IX N
آ خرَ E AE: KH AA R
آ خرَ E AE: KH AA R AA
آ خرَوُْن E AE: KH AA R UW N AE
آ خرَيِ نْ E AE: KH AA R IX: N AE

Journal of Theoretical and Applied Information Technology
31st July 2019. Vol.97. No 14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

آ خ رَ يِ نْ E AE: KH AA R IX: N
آ خ رَ E AE: KH AA R
آ خ ذِ ةَ E AE: KH IX DH AE T UH N
آ خ رِ E AE: KH IX R AA
آ خ رِ E AE: KH IX R
آذاَر E AE: DH AE: R
آ س يْ اَ E AE: S Y AE:
آ س يْ ا ن E AE: S Y AE: N

8. SPEECH RECOGNITION PATTERN MATCHING

After the feature is extracted from the spoken word, a matching algorithm is used to classify each word correctly. There are several classification algorithms [22] such as dynamic time warping (DTW) that measure in similarity between two sequences which may vary in time or speed, hidden Markov model(HMM) a collection of prototypical speech patterns are stored as reference patterns which represents the dictionary of candidate words. An unknown spoken utterance is matched with each of these reference templates and a category of the best matching pattern is selected, neural network (NN), and support vector machine(SVM) is one of the powerful state-of-the art classifiers for pattern recognition which uses a discriminative approach. Optimized margin, between the samples and the classifier border, helps to generalize unseen patterns. SVMs use linear and nonlinear separating hyper-planes for data classification, and hybrid approach. HMMs provide an elegant statistical framework for modeling speech patterns using a Markov process that can be represented as a set of states and a transition between these states that describes how the process change its characteristics in time. The HMM method provides a natural and highly reliable way of recognizing speech for a wide range of applications [23].

9. SPEECH RECOGNITION TOOLKITS

There are several speech recognition researchers used them in creation their system like

HTK: it stands for Hidden Markov Toolkit (HTK), it is an open source tool written completely in ANSI C, and it is for constructing and employing hidden Markov models. It has been designed for English recognition [24].

SPHINX: it is a speech recognizer engine, written completely in Java programming language. It is a flexible framework for speech recognition [25].

SCARF: It is a software toolkit designed speech recognition that performs a segmental analysis, in which each word can defined by a segment, thus allowing a phoneme, multi-phone, syllable and word level recognition.[26].

ISIP: Stand for The Institute for Signal and Information Processing (ISIP),it is developed at Mississippi State University then it becomes an available engine. The toolkit contains a front-end, a decoder, and a training component [27].

Kaldi :it written in C++ programming language and released under the Apache License that is very nonrestrictive, so it is appropriate for a large number of users[28][29].

10. CONCLUSION

Arabic speech recognition systems face many challenges. There is several pronunciation variability in Arabic language. The phonetic dictionary has a great impact on the accuracy of speech recognition system. The accuracy of the pronunciation dictionary depends on the selection of the phoneme set and the set of rules used in the creation of it. Romanized phoneme set provide good results than Arabic phoneme, but it isn't natural, reading it is difficult, and consume a lot of time to prepare and generate the phonemes. MFCC and HHM are almost used in all speech recognition systems.

REFRENCE

[1] Gordon R., Ethnologue: Languages of the World, Texas: Dallas, SIL International, 2005.
[2] Elmahdy et al. used acoustic models trained with large MSA news broadcast speech corpus to work as multilingual or multi-accent models to decode colloquial Arabic(2009).

[3] Habash, N. Y. (2010). *Introduction to Arabic natural language processing*. USA: Morgan and Claypool Publishers.

[4]Abushariah. M. A. M., “TAMEEM V1.0: speakers and text independent Arabic automatic continuous speech recognizer,” International Journal of Speech Technology, vol. 20, no. 2, pp. 261–280, 2017

[5] El-Imam. Y. A, “An Unrestricted Vocabulary Arabic Speech Synthesis System”, IEEE Transactions on Acoustic, Speech, and Signal Processing, Vol. 37,No. 12, Dec. 1989, pp.1829-45

[6] Anusuya MA, Katti SK. Speech recognition by Machine: a review. Int J Comput Sci Inf Secur 2009.

[7] Rudnicky, Alexander, I ., et al, "Survey of Current Speech Technology", Communication of the ACM, vol. 37, No. 3, Mar., 1994, pp. 52-57.

[8] Davis. S. B., Mermelstein. P., " Comparison of parametric representation for monosyllabic word recognition in continu-onsly spoken sentences " . IEEE Trans. ASSP, Aug., 1980.

[9] Abd Almisreb .A, Abidin. A. F, and Md Tahir .N, “Comparison of speech features for Arabic phonemes recognition system based Malay speakers,” 2014, pp. 79–83.

[10]- Saha,G.,Chakraborty,S.and Senapati,.s" A New Silence Removal and Endpoint Detection Algorithm for Speech and Speaker Recognition Applications" Proceedings of the NCC 2005, Jan. 2005.

[11]- Markel. J., and Gray. A., Linear Prediction of Speech, Springer-Verlag, New York, USA, ISBN: 0-13-007444-6, 1976.

[12]- Picone, J.,"Signal Modeling Techniques In Speech Recognition", IEEE Proceedings, vol. 81, no. 9, pp. 1215-1247, September 1993

[13]-Davis. S. B., Mermelstein. P., " Comparison of parametric representation for monosyllabic word recognition in continu-onsly spoken sentences " . IEEE Trans. ASSP, Aug., 1980.

[14] Karpagavalli .S and Chandra .E, "A Review on Automatic Speech Recognition Architecture and Approaches," International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 9, pp. 393-404, 2016.

[15] Abushariah, M. A. M. (2012). Automatic Continuous Speech Recognition Based On Phonetically Rich and Balanced Arabic Speech Corpus. Ph.D. Thesis, University of Malaya, Malaysia.

[16]Mimer, B., Stuker, S., & Schultz, T. (2004). Flexible decision trees for grapheme based speech recognition. In Proceedings of the 15th conference elektronische sprach signal verarbeitung (ESSV), Cottbus, Germany, 2004.

[17]Elshafei-Ahmed, M. (1991). Toward an Arabic Textto-Speech System. The Arabian Journal of Science and
Engineering, 16(4B), 565–583.

[18]Algamdi, M., Almuhtasib, H., & Elshafei, M.. Arabic Phonological Rules. [King Saud University. Journal of Computer Sciences and Information, 16, 1–25. 2004

[19] Hyassat. H., Abu Zitar R. .Arabic Speech recognition using SPHINX engine Int J Speech Technol, 9 , pp. 133-150, 2006

[20] Labidi. M, Maraoui.M, M. Zrigui, "New birth of the arabic phonetic dictionary", Engineering & MIS (ICEMIS) International Conference on. IEEE, pp. 1-9, 2016.

[21]Ali. M, Elshafei. M, Al-Ghamdi. M, Al-Muhtaseb. H, and Al-Najjar. A, “Generation of Arabic phonetic dictionaries for speech recognition,” 2008, pp. 59–63.

[22] Suman K. Saksamudre, P.P. Shrishrimal, R.R. Deshmukh “A Review on Different Approaches for Speech Recognition System”, International Journal of Computer Applications, Vol115 – No. 22, April 2015.

[23] Rabiner, L. R.," a tutorial in hidden Markov models and selected application in Speech Recognition", IEEE Proceedings, vol. 77, no.2, February 1989.

[24] Young .S, Evermann .G, Gales .M,Hain. T, Kershaw .D, Liu .X, Moore .G, Odell .J, Ollason. D, Povey .D, Valtchev .V, and Woodland .P, The HTK Book (for version 3.4). Cambridge University Engineering Department, 2009.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[25] Walker .W, Lamere. P, Kwok. P, Raj. B, Singh. R, Gouvea. E, Wolf. P, and Woelfel.J, “Sphinx-4: A flexible open source framework for speech recognition,” Sun Microsystems Inc., Technical Report SML1 TR20040811, 2004.

[26] Georey. Z and Patrick . N, "SCARF: A segmental CRF speech recognition system," Tech. Rep., Microsoft Research, 2009.

[27] Huang. K and Picone. J, “Internet-Accessible Speech Recognition Technology,” presented at the IEEE Midwest Symposium on Circuits and Systems, Tulsa, Oklahoma, USA, August 2002.

[28] Povey. D, Ghoshal. A, Boulianne. G, Burget. L, Glembek. O,Goel .N, Hannemann. M, Motlicek P, Qian .Y,Schwarz. P, Silovsky. J, Stemmer. G, and Vesely. K. The Kaldi Speech Recognition Toolkit. In IEEE 2011 Workshop on Automatic Speech Recognition and Understanding. IEEE Signal Processing Society, 2011.

[29] Ali, A., Zhang, Y., Cardinal, P., Dahak, N., Vogel, S., Glass, J., 2014. A complete kaldi recipe for building arabic speech recognition systems, in: Spoken Language Technology Workshop (SLT), 2014 IEEE, pp. 525–529.