International Conference on Computer and Communication Engineering (ICCCE 2010), 11-13 May 2010, Kuala Lumpur, Malaysia

Development of An Arabic Text-To-Speech System

Mustafa Zeki, Othman O. Khalifa, A. W. Naji
Electrical and Computer Engineering Department
International Islamic University Malaysia
Kuala Lumpur, Malaysia
almihimdi@yahoo.com

***Abstract*—Research on Text-to-speech technology has received the interest of professional researchers in many languages which is a consequence of wide range of applications where Text-To-Speech is implemented. However, Arabic language, spoken by millions of people as an official language in 24 different countries, gained less attention compared with other languages despite the fact that it has a religious value for more than 1.6 billion Muslim worldwide. These facts exhibit the need for a high quality, small size, and completely free Arabic TTS with the ability of future improvements. The vowelized written text of Arabic language carries the pronunciation rules with limited exceptions, so rule-based system with an exception dictionary for words that fail with those letter-to-phoneme rules may be a much more reasonable approach. This paper is a development of a rule-based text- to- speech Hybrid synthesis system which is a combination formant and concatenation techniques with acceptable naturalness. The simulation results of the system shows good quality in handling word, phrase, and sentence level compared to other available Arabic TTS systems. The accuracy of the overall system is 96%. Further improvements need to be done for stressed syllable position and intonation.**

***Keywords-component*; rule-based text-to-speech (TTS); speech synthesis; Arabic phonology; hybrid synthesis.**

I. INTRODUCTION

Speech is the primary means of communication between people. Speech synthesis is an automatic generation of speech waveforms. The dream of producing a talking machine started at the 18th century while recent progress in speech synthesis has produced synthesizers with high intelligibility but the sound quality and naturalness still remain a major problem. This fact makes speech synthesis an important field for investigation and improvement for the major languages including Arabic.

Speech synthesizer or as it known Text-to-Speech system is one of the important technology in the current time due to the expanding field of applications, it is used in multimedia applications to read e-mail, mobile messages, or in any kind of human-machine interaction. It is helpful and common among visually impaired people as a simple reading machine, gives the deafened and vocally handicapped an opportunity to communicate with people who do not understand the sign language. It can be used also in many educational tasks like spelling and pronunciation teaching aid for different languages. Arabic is the fourth most spoken language in our world with more than 442 million speaker spread in 23 countries as an official language [1]. Furthermore it carries a religious value for more than 1.6 billion Muslim according to [2].

The number of blinds in the Arab World is around 5 millions living in a population around 340 million people [3]. So, it's an important issue to build Arabic TTS which is reliable, intelligent and user friendly system to give those people a chance to use the technologies like text messages, emails, and web sites using their native language.

The text-to-speech (TTS) synthesis procedure consists of two main phases. The first one is text analysis, where the input text is transcribed into a phonetic or some other linguistic representation, and the second one is the generation of speech waveforms, where the acoustic output is produced from this phonetic and prosodic information. The main techniques used in speech synthesis design are Articulator synthesis, Formant synthesis, and Concatenative synthesis. Articulatory synthesis attempts to model the human speech production system directly. Formant synthesis, which models the pole frequencies of speech signal or transfer function of vocal tract based on source-filter-model. Concatenative synthesis, which uses different length pre-recorded samples derived from natural speech [4].

In theory, the most accurate method is articulatory synthesis which models the human speech production system directly, but it is also the most difficult approach. So, the available TTS systems mostly use either concatenative or formant synthesis technique. Each technique has its own points of strength and weakness and suits a specific language while doesn't for others [5],[6].

An interesting approach is to use a hybrid system where the formant and concatenative methods have been applied in parallel to phonemes where they are the most suitable [7]. In general, combining the best parts of the basic methods is a good idea, but in practice, controlling of synthesizer may become difficult [5].

II. ARABIC PHONETIC SYSTEM

The phonetic system of Modern Standard Arabic has basically 34 phonemes, which consists of 26 consonants, 3 short vowels, 3 long vowels and 2 semivowels [8]. Arabic uses another diacritic called *sukun*. A *sukun* implies a consonant not followed by a vowel; i.e. implies a cluster of consonants, so it's not necessary to write *sukun* in the input text because each letter without a short vowel is considered by default with *sukun*.

978-1-4244-6235-3/10/$26.00 ©2010 IEEE

TABLE I. ARABIC CONSONANTS AND THEIR PHONEMIC TRNSCRIPTION

		Bilabial	Labiodental	Interdental	Alveodental	Palatal	Velar	Uvular	pharyngeal	laryngeal
Stops	Voiced	b/ب			d/د d2/ض	dZ/ج				
	unvoiced				t/ت t2/ط		k/ك	q/ق		?/ء،ا
Fricative	Voiced			D/ذ D2/ظ	z/ز			G/غ	?#/ع	
	unvoiced		f/ف	T/ث	s/س s2/ص	S/ش	x/خ		H/ح	h/هـ
Nasal	Voiced	m/م			n/ن					
Trill	Voiced				R/ر					
Lateral	Voiced				l/ل					
Semi vowels	Voiced	w/و				j/ي				

The Arabic phonetic system differs from the Latin ones essentially by emphatic and glottal phonemes. Table I shows the used phonetic transcription for the Arabic consonants and their equivalents.

A. Syllables

The syllabic structures in Arabic are limited in number and easily detectable. Unlike other languages, every syllable in Arabic begins with a consonant followed by a vowel which is called the nucleus of the syllable. Short vowels are denoted by (V) and long vowels are denoted by (*VV*). It is obvious that the vowel exists in the second place of the syllable. These features facilitate the process of syllabification.

Arabic syllables can be classified either according to the length of the syllable or according to the end of the syllable**.** Short syllable occur only in CV form, because it is ending with a vowel so it is open. Medium syllable can be in the form of open CVV, or closed CVC. Long syllable has two closed forms CVVC, and CVCC.

Arabic words contains at least one syllable, most contain two or more syllables. The longest word is combined of five syllables. Table II illustrates Arabic syllables. Some of the Arabic words are spelled together forming new long words with 6 syllables like (سألتمونيها), or 7 syllables like (أتلزمكموها).

TABLE II. ARABIC SYLLABLES TYPES

Syllable type	Arabic example		English meaning
CV	بـ	/bi/	in, at
CVV	ما	/ma2:/	What
CVC	مِن	/min/	From
CVCC	حَرْب	/Ha2Rb/	War
CVVC	نار	/na2:R/	Fire
CVVCC	سارّ	/sa2:R1R/	Delightful

III. SYSTEM DESCRIPTION

The complete design of the proposed synthesis system is shown in Fig. 1. The Text-To-Speech system includes preprocessing stage for text normalization. Natural language processing (NLP) decides the stress pattern and converts the letters to its equivalent phoneme representation. DSP stage produces speech out of the resultant phonemes.

A. Preprocessing Stage

Preprocessing phase is required to prepare the raw text to be suitable to enter the processing stage. It detects the words of each sentence, spaces between them, punctuation marks, or any other non-Arabic symbols.

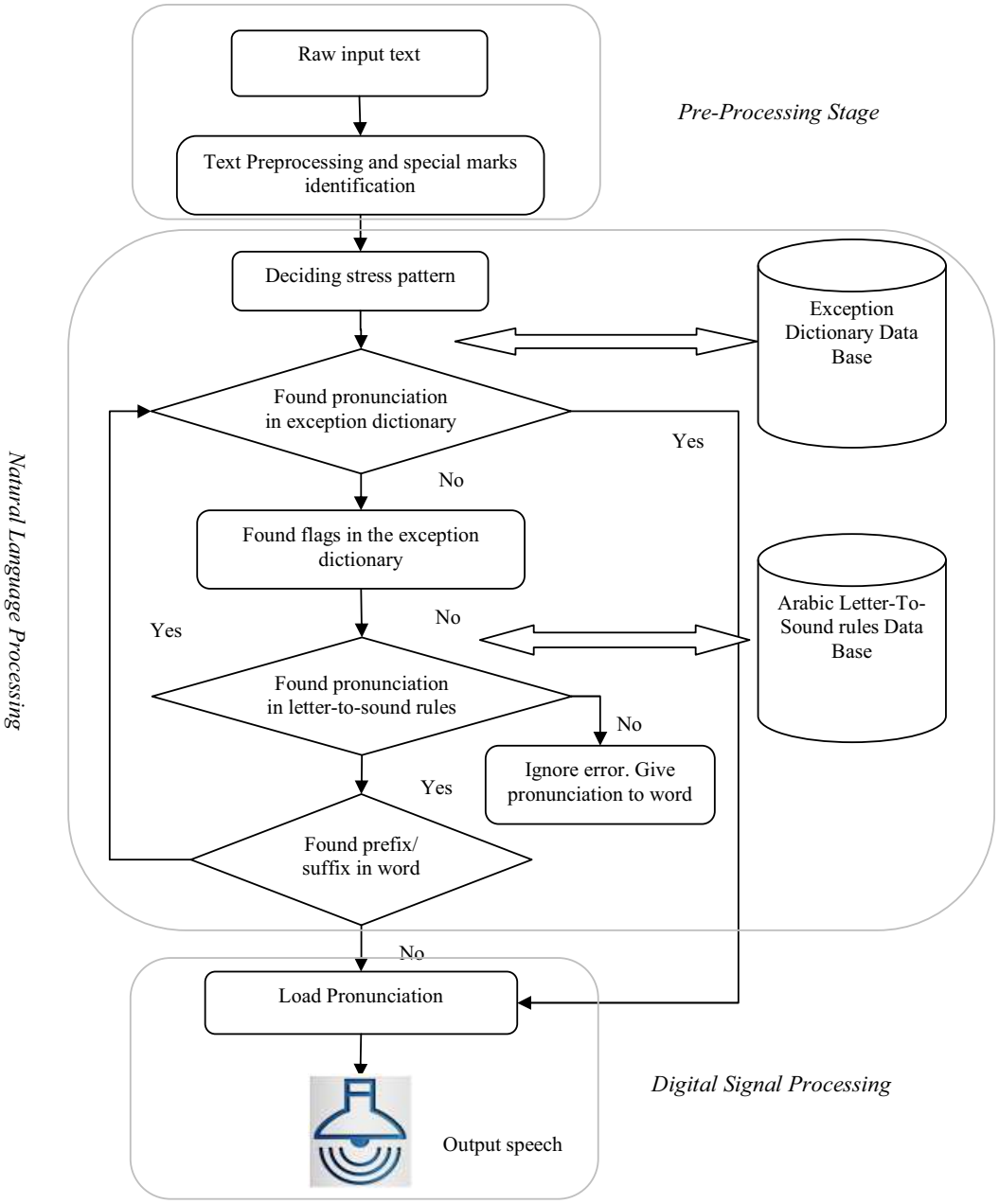

Figure 1. Arabic Synthesis System Structure

B. Natural Language Processing

This phase relies on the developed data base to map each word to its exact phonemic representation equivalent. The data base is formed of three main parts:

1) Exceptions Dictionary: It contains list of words whose pronunciations are given explicitly, rather than determined by the Pronunciation Rules. If the Pronunciation rules are applied to a word and indicate a standard prefix or suffix, then the word is again looked up in Exception Dictionary List after the prefix or suffix has been removed. Rather than a full pronunciation, just the stress flag may be given, to change where it would be otherwise placed by the Pronunciation Rules. A pronunciation may also be specified for a group of words, when these appear together. Up to four words may be given, enclosed in brackets. This may be used for change the pronunciation or stress pattern when these words occur together, run them together, pronounced as a single word, or to give them a flag when they occur together. Common set of words mostly need no vowelization (*Tashkeel*)

2) Arabic Pronunciation Rules (letter-to-sound): This part of the data base specifies the phonemes which are used to pronounce each letter, or sequence of letters. Some rules only apply when the letter or letters are preceded by, or followed by, other specified letters [9], [10]. To find the pronunciation of a word, the rules are searched and any which match the letters in the word will be given a score depending on how many letters are matched. The pronunciation from the best matching rule is chosen. The pointer into the source word is then advanced past those letters which have been matched and the process is repeated until all the letters of the word have been processed. It is important to mention that the Exception Dictionary is searched first while it contains the words with special pronunciation.

3) Arabic phonemes: Defines all the phonemes which are used by Arabic language, together with their properties and the data for their production as sounds. Vowels are generated based on formant synthesis, while special Arabic consonants like: (ض، ق، ط، ظ، ص،) are pre-recorded wav files based on concatenation synthesis.

Arabic words that are not found in the exception dictionary with their stress flag will follow a special pattern for syllable lexical stress. It is designed to apply stress according to a set of rules. Syllable has three lexical stress levels: primary (1), secondary (2), and weak or no stress (3), as reported in [7] and [11] the rules that determine the stress are:

- When a word is made up of a string of the CV type syllables, the first syllable receives the primary stress and the remaining syllables receive no stresses, e.g. (كَتَبَ) : كَ(1)تَ(3)بَ(3) : CV(1)CV(3)CV(3) "he wrote"
- When a word contains only one long syllable, the long syllable receives the primary stress and the rest of the syllables go unmarked, receiving no stresses. The final long syllable never receives a primary stress. (كاتِب) : كا(1)تِب(3) CVV(1)CVC(3) "writer"
- For polysyllabic word, a stress is placed on the first long syllable counting from the penultimate. The nearest long syllable to the beginning of the word receives the secondary stress. (مُدَّخَراتُهُم) : مُد(3)دَ(2)خَ(3)را(1)تُ(3)هُ(3)م(3)
CVC(3)CV(2)CV(3)CVV(1)CV(3)CVC(3) "their savings"

C. Digital Signal Processing

In this part of the synthesis system the resultant phonemic representation of the input text with the special stress and pause flags will be transformed into the proper utterance and can be saves as a wave file. Table III shows some Arabic words and their resultant phonemic representation and stress flags taken from our system.

IV. Testing and Results

The simulation results shows 96% accuracy for the phoneme representation (Letter-To-sound), all the diacritics of the standard Arabic script including were successfully mapped into their equivalent pronunciation rules. The words with special pronunciation were listed in the exception data base, since that data base is checked before the rules data base, the system ensures error free phoneme representation. In order to assess the quality of our output speech, a subjective test was performed. A set of 25 phonetically balanced sentences was used as the test material. The test sets were played to ten volunteer listeners who were asked to rate the system intelligibility, naturalness and overall voice quality on a scale of 1 to 5. The volunteer listeners were asked to give the rating based on how good they thought the sentences were without any further definition of the word "good". The listeners were encouraged to use the full five point scale. The average scores obtained are 4.2, 3.7, and 3.8 for intelligibility, naturalness and overall voice quality respectively.

TABLE III. Phonemic representation of Arabic words

English	Arabic	IPA	Proposed System
Hello	مرحبا	/marħaban/	/m'a2RHa2ba2n/
Goodbye	مع السلامة	/maʕ assala:mah/	/ma2?a2s1sa2l'a2:ma2/
House	بيت	/bai:t/	/ba2jt/
Thank you	شكرا لك	/ʃukran lak/	/S'UkRa2n11la2k/
Morning	صباح	/ sˤaba:ħ/	/s2a2ba2:H/
Noon	ظهر	/ðˤuhr/	/D2Uhr/

V. Conclusion

A rule-based hybrid synthesis Arabic TTS system was developed. Phonemes were the essential elements of the synthesizer, our Arabic TTS system is vocabulary independent with intelligible output speech, so it can handle all types of input text. It has the flexibility of changing the speaker from male to female and other sound variants like whispering. The standard Arabic text is mostly unvowelized; hence the need of vowelization (*Tashkeel*), our system omits the need to some of the vowelization symbols like *sukoon* and has the ability to enrich the exception dictionary by listing the exact pronunciation of the common words, here, no need to vowelize them. Comparing with other available Arabic TTS systems, our Hybrid TTS has small size, high accuracy, and vocabulary independence features which make it in general more reliable than other TTS systems. The system is free for distribution and for development.

References

[1] Bateson, Mary Catherine, Arabic Lanuage Handbook. Georgetown University, 2003.

[2] Pew Research Center, "Mapping the Global Muslim Population," The Pew Forum on Religion & Public Life, 2009.

[3] Regional Office of The Middle East, "The Annual Report for the Regional Director," World Health Organization.

[4] R. D. Rodman, Computer Speech Technology. London: Artech House Publishers, 1999.

[5] S. Lemmetty, Review of Speech Synthesis Technology, Master Thesis. Helsinki University of Technology, 1999.

[6] M. G. Fatima Chouireb, "Towards a high quality Arabic speech synthesis system based on neural networks and residual excited vocal tract model," Springer, Oct. 2008.

[7] G. Fries, "Hybrid time- and frequency-domain speech synthesis with extended glottal source generation," in IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 1, 1994, pp. 581-584.

[8] S. Al Ani, "Arabic Phonology: An Acoustical and Physiological Investigation," The Hague, 1970.

[9] M. E. M. Al-ghamdi, "Phonetic Rules in Arabic Script," King Fahd University of Petroleum & Minerals, 2002.

[10] I. Noriddeen, Linguistic phonetics, 1st ed. Lebonan, 1992.

[11] Z. Zemirli, "ARAB_TTS: An Arabic Text To Speech Synthesis," in IEEE International Conference on Computer Systems and Applications, 2006, pp. 976-979.