TOWARD AN ARABIC TEXT-TO-SPEECH SYSTEM

M. Elshafei Ahmed*

Department of Systems Engineering
King Fahd University of Petroleum & Minerals
Dhahran, Saudi Arabia

الخلاصـــة :

تحتوى هذه المقالة على وصف لنظام تجريبي لخوارزمي غايته نطق حروف الكتابة العربية باستعمال المقاطع الصوتية . وهذا النظام يستخدم فئـة شاملة من الألفونيات العربية لتوليد الكلام طبقا للغة العربية الحديثة والفصحى على السواء . وتحتوى هذه الفئـة من الألفونيات على حوالي (١٥٠) ألفون عربياً مع بعض الثنائيات من الأصوات الصامتـة واللينة لتسهيل نطق الكلام بمساندة الحاسب . كما تصف هذه المقالة عدداً كبيراً من قواعد النطق طبقا لهذه الفئـة من الألفونيات . وتغطى هذه القواعد تتابع الاصوات المصمته مع اللينة واستعمال الحركات القصيرة والطويلة وتأثير الأصوات المفخمة والحلقية على مخارج الأصوات ، وكذلك قواعد نطق أداة التعريف وقواعد الوقف والتنوين والـغُـنه والإدغام والإقلاب والإخفاء وكذلك قواعد الحاق الكلمات ببعضها .

*Address for correspondence:
KFUPM Box 405
King Fahd University of Petroleum & Minerals
Dhahran 31261
Saudi Arabia

ABSTRACT

This paper describes an experimental system for algorithmic based segmental Arabic text-to-speech system. The system utilizes a comprehensive set of Arabic allophones for speech synthesis in accordance with both Modern Standard Arabic and the classical Arabic language. The proposed allophone set contains about 150 Arabic allophones and vowel/consonant combinations to simplify computer voice production. The paper also describes an extensive set of Arabic letter-to-sound rules from voweled text (with full diacritics) based on this allophone set. This set of rules governs consonant/vowel combinations, usage of short and long vowels, the coarticulation effects of emphatics and pharyngeals, pronunciation of the particularization, rules for word final aspiration, *Tanween* and *Ghonna* rules, rules for conversion, suppression, and assimilation, and rules for combining words.

TOWARD AN ARABIC TEXT-TO-SPEECH SYSTEM

1. INTRODUCTION

The last few years have witnessed an increasing interest in speech processing of the Arabic language [1–12]. However, a great deal of research still needs to be done in order to realize automated text to speech systems with comparable quality to the available commercial and laboratory systems for the English language. The progress in the English text-to-synthetic speech systems has been made possible by advances in linguistic theory, acoustic phonetics, computer characterization of English sound patterns, perceptual psychology, mathematical modeling of speech production, in addition to advances in computer hardware.

There are currently several approaches for Arabic speech synthesis; the direct storage of words and sentences [3, 23], the syllable method [5], sub-syllable method [5, 6], the diphone method [7], and the allophone method [1–3]. These approaches vary in complexity, memory requirements, and speech quality. In this paper we describe an algorithmic method for allophone synthesis of Arabic speech. The main advantages of this approach are the small memory requirement, the small number of the basic sound units required to synthesize an arbitrary Arabic word, and the simplicity of the hardware. However, the use of a limited number of fixed sound units may cause the resulting speech to sound "mechanical". The direct concatenation of the allophones could cause unnatural abrupt transitions and neglects the coarticulation and interaction between consonants. To overcome this problem, the paper proposes over 150 allophone and consonant/vowel combinations to cover most of the prominent consonant interactions. Model parameter interpolation was also implemented over 2–5 msec to provide a smooth transition between the allophones. At a later stage, heuristic rules [29] will be developed to smooth the transition on a case-by-case bases.

An overview of a text-to-speech system is shown in Figure 1. The presence of the diacritic marks in the Arabic text is essential for the implementation of the automatic text-to-speech system. Unfortunately, most modern written Arabic, as in books and newspapers, is at the best partially vowelized. Hence a preprocessor for automatic vowelization of the text must be implemented before applying the text-to-speech rules. The automatic vowelization generator requires integration of morphological, syntactical, and semantic information [13–18, 23].

The text with full diacritics is then passed through a second preprocessing step wherein a lexicon is used to replace numerals, abbreviations, special symbols as %, +, −, =, &, $, *etc.*, and a few exceptions as يس *Yassen*, هذا ، كجم *etc.*, by an equivalent phonemic spelling.

Next, the raw text produced from the preprocessing steps is partially parsed into breathing groups. This can be accomplished approximately in many ways. A simple rule that works most of the time as a natural speaker is obtained by locating the breathing group boundary at the following locations [26] whichever is encountered first: (1) at a punctuation mark; (2) preceding a conjunction; (3) before a noun phrase or a verbal phrase; (4) before a prepositional phrase; (5) after a fixed number of characters have appeared in the input.

In the next step, the text is processed by the first level of letter-to-sound rules. The objective here is to produce an abstract phonetic transcription of how the text is actually spoken. At this level the speech is also parsed into syllabic units, for example:

مَن رأى العَنْبَر is pronounced مَر - رَ - ءَل - عَم - بَر .

The next level is the translation of the abstract phonetic transcript to an allophonic transcription, wherein each phoneme sound is replaced by a context dependent acoustic version of the phoneme [26, 29]. In the mean time a companion description of the stress and inflection patterns of speech are generated [23, 29].

Finally, a parametric description of the articulation model, or the synthesis model, is used for generation of the speech sound. The parameters are modulated by the prosodic information as well as other heuristic rules for graded changes in the synthesis model parameters [29]. Several synthesis models are presently known:

1. Formant Synthesizer [29, 30, 32].
2. LPC Synthesizer [1, 2, 31].
3. Channel Synthesizer [30].
4. Articulatory Synthesizer [28, 30].

The synthesis model should allow the synthesized speech to be modified in fundamental frequency,

Figure 1.

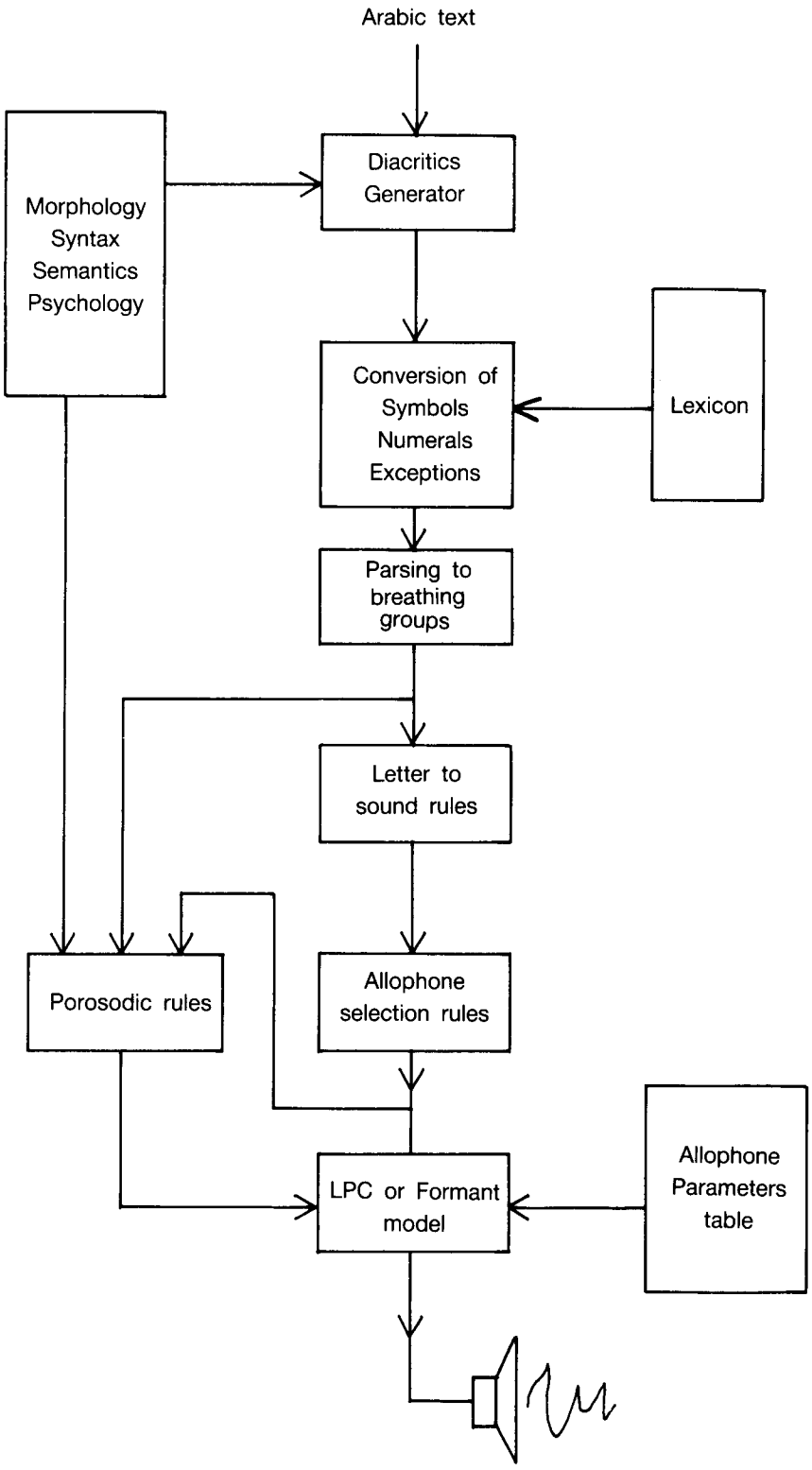

Figure 1.

amplitude, and duration, as well as performing some sort of parameter smoothing at boundaries between waveform pieces.

In this paper we describe an experimental system for rule based segmental text-to-speech system. Specifically, the paper addresses the following issues:

1. proposes an allophone set for speech synthesis;
2. presents a set of rules for conversion of fully diacritized text to an abstract phonematic transcription;
3. provides an extensive set of rules for conversion of the abstract phonematic transcription to allophone strings.

The development work was carried out on a commercial speech editing laboratory. The synthesis model is based on linear prediction coding [1, 31] using 10 filter coefficients and pitch synchronous analysis/synthesis. A detailed description of the model was given in [1]. Although the model was flexible enough to allow changing of duration, stress, and fundamental frequency, the speech quality produced by this method was still unsatisfactory for many applications because of the difficulty in smoothing of the filter coefficients without disturbing the formants trajectories, and because of the limited source excitations model, in addition to the limited bandwidth (4 kHz) which causes difficulty in pronunciation of fricative sounds. A formant model [32] is currently being investigated and recommended for future investigations. Nevertheless, speech synthesis by concatenating strings of Arabic allophones was displayed in [2]. The early version of the text-to-allophone program was also demonstrated separately in [4]. The main advantage of the allophone method is its simplicity and small memory requirements.

In the next section we introduce the proposed allophone set and in Section 3 an extensive set of Arabic pronunciation rules will be presented.

2. ARABIC ALLOPHONES

It is impressive to mention here the first phonetic study of the Arabic language (probably the first published study in phonology of all) of Ibn Sina (370−428 A.H.) [21]. Ibn Sina's treatise explains the mechanism of production of each letter as well as the points of articulation with reference to a fairly accurate anatomy of the vocal tract.

Among the recent studies is [19] which presents a comprehensive introduction to the modern classification of the Arabic sounds, their points of articulation, as well as some of the Arabic allophones. Al Ani [10] investigated the modern dialectal Arabic of Iraq using spectrographic techniques. El-Emam and Dannan [8] studied the allophones of the "Modern Standard Arabic MSA". In [2] about 100 basic allophones of classical Arabic were isolated and used for allophone-based speech synthesis system.

Table 1 shows the Arabic letters and their equivalent International Phonemic Alphabet IPA [3, 9].

The proposed Arabic allophone set contains about 150 allophones and consonant/vowel combinations to simplify computer voice production. Table 2 shows the Arabic set combined with a set of English allophones [25] for comparison, as well as to give the system some ability to pronounce English words, abbreviation, and names within the Arabic text.

The investigated set of Arabic allophones incorporates a number of allophones that appear in MSA [8], the basic allophones that characterize the classical Arabic and Quranic recitation [1, 2], as well as some vowel/consonant clusters to simplify computer speech synthesis.

The proposed allophone set is capable of producing synthetic speech for Quran, classical Arabic, and Modern Arabic with a relatively high degree of naturalness. However, it should be emphasized here that it is not the purpose of this study to define a minimal set of Arabic allophones from the linguistic point of view, but rather to identify a group of speech segments for computer generation of synthetic speech with relatively high naturalness.

The notations used to label the newly introduced allophones follow those used in the literature to describe the English allophones, see for example [26].

The allophones were isolated and digitally encoded and verified by inspection of the formant trajectories, as well as by using analysis-by-synthesis technique.

The Arabic language is characterized by a relatively large number of back consonants (produced in the back of the oral cavity). The back consonants cause a complex coarticulation phenomenon in Arabic speech. The back consonants that characterize the Arabic language are the pharyngealized set shown in Table 1. If we take another careful look at these consonants we can divide them further into the following groups; the emphatic consonants such as *Dhad* /D/, *Sad* /S/, *Tah* /T/, and *Zah* /ð/; the pharyn-

Table 1. Arabic Consonants.

			Bilibial	Libio-dental	Inter-dental	Alveo-dental Alveolar	Palatal	Velar	Oropha-ryngeal Uvular	Pharyn-geal	Glottal
Stops	Voiced	Pharyngealized				D ض					
			b ب			d د					
	Unvoiced	Pharyngealized				T ط			q ق		
						t ت		k ك			E ء
Fricative	Voiced	Pharyngealized			ð ظ				ɣ غ		
					ð ذ	z ز				? ع	
	Unvoiced	Pharyngealized				S ص			x خ		
				f ف	Θ ث	s س	ʃ ش			H ح	h ه
Affricative							dz ج				
Nasals	Voiced		m م			n ن					
Resonants	Voiced	Pharyngealized				L ل ر					
			W و			l ل r ر	y ى				

geals such as *Ghain* /ɣ/, *Khah* /x/, and *Qaf* /q/; and semi-emphatic such as *Lam* /L/ in *Allah* and *Raa* /r/. The emphatic and semi-emphatic consonants not only cause the vowel after them to be pharyngealized but also they influence the vowels preceding them and some times the entire neighboring segments [8–10]. This phenomenon is similar to the R-controlled effect in English pronunciation. In the allophone-based English speech synthesis, the R-controlled vowels are listed separately as clusters of vowels/R. Accordingly, in this study, we also propose a set of vowel/consonant clusters to simplify the Arabic speech synthesis of the emphatic consonants.

Before we describe the allophone set, let us define the following groups of letters:

1. Group A: the Emphatic Consonants: /T/, /S/, /D/, and /ð/.
2. Group B: the Pharyngeals: /q/, /x/, and /ɣ/; and /r/.
3. Group C: the *Madd* letters: *Alif*, "ا", *Ya'a*, "ى", *Waw*. "و".
4. Group D: the rest of the letters (except the pharyngealized *Lam* /L/).
5. Group E: Glottal/Pharyngeals (*Al-Ezhar* letters); /E/, /h/, /H/, /?/, /x/, /ɣ/.
6. Group F: *Ash-Shamsi* letters: /t/, /Θ/, /d/, /ð/, /z/, /s/, /ʃ/, /S/, /D/, /T/, /ð/, /l/, /n/.
7. Group G: *Al-Qamari* letters: /E/, /b/, /dz/, /H/, /x/, /?/, /ɣ/, /f/, /q/, /k/, /m/, /w/, /h/, /y/.
8. Group H: *Muqalqal* letters (aspirated): /q/, /T/, /b/, /dz/, /d/.
9. Group I: *Ikhfa'a* letters: /t/, /Θ/, /s/, /ʃ/, /dz/, /d/, /ð/, /z/, /S/, /D/, /T/, /ð/, /f/, /k/, /q/.
10. Group J: Voiceless Fricative consonants: /f/, /Θ/, /s/, /ʃ/, /h/, /H/, /S/, /x/.
11. Group K: Stops: /D/, /d/, /t/, /T/, /k/, /q/, /b/.
12. Group L: the Consonants: /dz/, /q/, /k/.

Table 2.

Allo-phone	English Usage	Arabic Usage
Pauses		
/PA1/	before V-stops and Fricatives	(في التشديد)
/PA2/	same as /PA1/	(بين المقاطع)
/PA3/	between words, before voiceless stops	(بين المقاطع، قبل حروف العطف، بين الكلمات)
/PA4/	between clauses, sentences	(بين الكلمات والجُمل)
/PA5/	same as PA4	(بين الجُمل)
Short Vowels		
/AE/	hAt, extrAct	(فَتحة) بَ، تَ، ثَ
/IH/	sIt, strandEd	(كسرة) بِ، تِ، سِ
/UH/*	fUll, lOOk	(ضمة) بُ، سُ، عُ
/AA/*	cAr, cOtton	رَ، غَ، خَ
/EH/	End, extEnd	بِنت، بناتِه
/AO/	sOng, tAlk	
/AX/	sUcceed, instrUct	
/An/*		مَن، نَم
/In/*		مِن، مِمّا
/Un/*		نُنْجي
/IX/*		صِنْف، صِل
/UX/*		عُصن، صُن
/AA1/*		فَط، بَط، بَـض
/AH/		ضَحل، قَد، قَهقَه
Long Vowels		
/AE:/		تَـاب، بَـاب
/IY/	See, pEOple	فِيل، حَبِيب
/EY/	grEAt	
/AY/*	skY, trY, guY	غَيْـس، خَير، غَيْب
/AY1/*		زَايْـف، سَيْـف
/AY2/*		ضَيف، طَير
/AA:/		خال، رَاح
/OY/*	bOY, tOY	
/UW1/*	compUter	طيور
/UW2/	tWO, fOOd	دُون، تُو بُو ا
/OW/	gOne, snOW	قُوا

Allo-phone	English Usage	Arabic Usage
/AW/*	sOUnd, mOUse	
/AW1/*		دَوْما، نَوم، لَـوْ
/AH:/		ضَال، طَال، ضَاق
/An:/		إمَـام، إيـمَـان
/Un:/		نُـون
/IX:/		قِـيص، بـصِـيص
Voiced Fricatives		
/VV/*	Vast, ValVe	
/DH1/	They, This	
/DH2/	word final, between vowels	
/ZZ/*	Zoo, phaSe	زَ، زِ، زُ
/ZZ1/*		زَاع، فـز
/ZH/*	beiGe, pleaSure (MSA)	جَ، جِ، جُ
/ZH2/*		أذكار
/ZH3/		ذَ، ذِ، ذُ
/ZH4/*		فَـظْ، أظْلم
/ZH5/		ظَ، ظِ، ظُ
/GH1/*		صاغْ، إغْـتصب
/GH2/		غَ، غِ، غُ
/AN1/*		يعْمل
/AN2/		عَ، عِ، عُ
/AN3/		عَـظيم، عَـطس
Voiceless Fricatives		
/FF/*	Fast, Foot	فَ، فِ، فُ، فَـات
/FF1/*	Far	فَـضْـل، فَـقَّه
/TH/*	tooTH, THin	ثَ، ثِ، ثُ، ثِـمار
/SS/*	teSt, See	سَ، سِ، سُ
/SS1/*		سَـطر، قِرطاسْ
/S1/*		صَ، صِ، صُ، صْ، صَـاح
/SH/*	fiSH, SHip	شَ، شِ، شُ، شَـعر، شِمال
/SH1/*		قشـطْ، طاش، شَـطّ
/HH1/*	Hat, He	هَـ، هِـ، هذا، هِنْدُ
/HH2/	Horn, Hope	هُو، هُـنا
/H1/*		حَ، ح

Allo-phone	English Usage	Arabic Usage
/H2/		حُ
/H3/*		طَاحْ ، صَاحْ ، حَقّ ، حَضّ
/KH/*		خَ ، خِ ، خُ
/WH/	WHig, White	
Affricatives		
/CH/	CHurch, CHina	
/JH/	juDGe, inJured	جَ ، جِ ، جُ
/JH1/		(مشددة) فُجّار
/JH2/		فَـج
Nasals		
/MM/*	Milk, Man	مَ ، مِ ، مُ
/MM:/		(غنه) أمّـا ، مِـمّـا ، كـمّـن
/M1/*		مَضرب ، قامْ ، ضَـمْ
/NM/		(صوت بين الميم والنون) من ما ، عنبر
/NN1/*	thiN, earN	نَ ، نِ ، نَـم ، مَـن
/NN2/	No, Normal	نُـطـق ، نُصح
/NG/	aNger, straNger	ساكنه قبل ك ، ق ، ج ، أنـقَرة ، إنكسر
/NN3/		نُـور
/NN4/		نون ساكنه قبل مجموعة I مُـنـتهى مَـن تاب
/N1/*		(مفخمه) نَصر ، صَان
/NN1:/		إنّ مِنّـا
/N1:/		تصـنّـط
/NY/		من يكُن ، من يعمل
Resonants		
/WW/	We, Warrant	وَ ، وِ ، وَجَـدَ
/WW:/		لَـووّ
/WW1/		وغـفِر ، وَقـرْ
/RR1/	Read, Ray	رِ ، ربا (غير مفخمه)
/RR2/	bRown, Crane	أمـرْ ، قَـفـر
/RR3/		رُ ، رَ ، رَب ، رَاح (مفخمه)
/LL/*	Like, heLLo	لَ ، لِ ، لُ ، لْ
/LL1/*		اللّـهُ ، صلّى اللّـهُ
/EL/	littLE, angLE	إل ، فى المدينه بالقطار
/YY1/	compUter, cUte	أيُّـــها
/YY2/	Yes, YoYo	ي ، يُوجد ، يَمْنع
/YY3/		يَط بُـع ، يَـظ بط

Allo-phone	English Usage	Arabic Usage
Voiced Stops		
/BB1/	riB	أبَـــا ، إبن
/BB2/	Beast, Bee	بَ ، بِ ، بُ ، باع ، بناء ، بُـشرى
/BB3/		(ساكنة مقلقه) تَـبْ ، دُبْ ، ابْـتر
/B1/		ضَبت ، قِبـط ، صَـبْـرْ
/B2/		بَصرة ، بَطة ، بِيُـصّ
/B3/		(ساكن نهائي) صَـبْ ، طابْ ، غابْ
/DD1/	enD, coulD	عِـدْها
/DD2/	Down, Drain	دَ ، دِ ، دُ
/DD3/		(مقلقله) تَـعـدْ ، أحـدْ ، أدبَر
/DD4/		قَـدْ ، صَـدْ
/DA1/		ضَ
/DA2/		ضِ
/DA3/		ضُ
/DA4/		إضح ، غَـضّـة
/DA5/		(ساكن نهائية) خاض ، أضْـرَمَ
/GG1/	Guest, Gift	
/GG2/	Got, Goat	
/GG3/	peG, anGer	
Voiceless Stops		
/PP/	Pay, triP	
/TT1/	tesTs, iTs	فـتـنه
/TT2/	streeT, Two	تَ ، تِ ، تُ ، نبات
/TT3/		تَصهروا ، تَـضحك
/TT4/		فَـتّـت
/T1/		طَ
/T2/		طِ
/T3/		طُ
/T4/		طْ قطن
/T5/		(مقلقلة) قط ، أحاط
/KK1/	Cute, Clown	لَ ، لِ ، كَلب ، كِتاب
/KK2/	tasK, speaK	فك ، دَكْ ، شكْـوَة
/KK3/	Quick	لِـهُ
/KK4/		صلْ
/KK5/		كـظْماء
/QQ1/		فَـقْـد
/QQ2/		قَ

Allo-phone	English Usage	Arabic Usage
/QQ3/		قِ
/QQ4/		قُ
/QQ5/		(ساكنة مقلقلة) أفاق ، أقفَل
/EE1/		(نهائية) ضَوءْ ، سماءْ
/EE2/		طاطأ
/EE3/		آمين
/EE4/		إنَّ
/EE5/		أُحُدْ
R-Controlled Vowels		
/ER1/	lettER	
/ER2/	bURn, bIRd	
/OR/	fORtune, stORe	
/AR/	alARm, fARm	ضارْ ، غارْ
/YR/	clEAR, hEAR	بِير ، عِيرْ
/XR/	hAIR, repAIR	
/UR/		زُور ، نُورْ
/AR1/		دارْ ، عارْ
/IR1/		بِركة ، تِرْكة
/IR2/		طِرْ ، قِرْ
Pharyngealed Controlled		
/IS/		حِصّه ، قِصط
/US/		فُصحه
/AS/		قاص
/aS/		فَصْبر ، بَصـــرة
/ID/		بِيض ، غِضروف
/UD/		مُضنى
/AD/		غَضبان
/aD/		وَاضرب
/IT/		إطفاء ، فِطرة
/UT/		لوط ، تُطبَع
/AT/		سَطر ، قَطرة
/aT/		تَطلَع
/IZ/		عِظهُم ، غِظهُم
/UZ/		مُظلم

Allo-phone	English Usage	Arabic Usage
/AZ/		قَظمُ
/aZ/		استَظلم
/IQ/		نِقمه
/UQ/		سُقناة ، مُقتبل
/AQ/		صَقرٌ
/aQ/		استقبَل ، مَقتَل
/AL/		صَلى الله
/AKH/		صَخِر
/aKH/		فَخر
/AGH/		صَغرٌ ، ضَغْطُ
/aGH/		يَغرَق ، استَغرق

* These allophones are actually Diphthongs.
* These allophones can be doubled.

2.1. Pauses

Short pauses appear in Arabic words before certain consonants, mainly to allow a transition period for the articulators to take their new positions especially in vowel-consonants sequences. Pauses also appear in some geminated utterances, between syllable patterns, and between words and sentences. /PA1/ and /PA2/ are usually used in geminated utterances like the voiced stop /BB/ in *Abba* أبّا .

There are five syllable patterns in Arabic; CV, CVV, CVC, CVVC, and CVCC, where C stands for a consonant, V stands for a short vowel, and VV stands for a long vowel. /PA2/ normally appears between such syllable patterns. /PA3/ and /PA4/ come between words. /PA3/ occurs before/after conjunctions as *Wa* وَ , and /PA4/ appears before conjunctions, while /PA4/, and /PA5/ appear between clauses and sentences.

2.2. Short Vowels

The regular Arabic short allophone vowels are /AE/, /IH/, and /UH/. The /AA/ is the short allophones of /a/ which usually appears after a pharyngealized letter of group *B* and after /l/ in *Allah* [19]. When a group *A* emphatic consonant is vowelized with /a/ the short vowel /AH/ may be used. /AA1/ is used when the unpharyngealized vowel /a/ precedes an unvoweled pharyngealized letter of groups *A* or *B*.

When a short vowel is located between two nasal sounds as /m/ and /n/, it is likely to be nasalized [8, 27]. The nasalized versions of /a/, /i/, and /u/ are /An/, /In/, and /Un/ respectively. Finally /IX/ and /UX/ are pharyngealized versions of the vowels /i/ and /u/ respectively.

2.3. Long Vowels

The regular Arabic long allophones are /AE:/, /IY/, and /UW2/. The length of a long vowel is normally equal to two short vowels. In general, longer vowel allophones can be constructed by multiple of short allophones. /AA:/ is the long allophone of /a/ which usually appears after a pharyngealized letter of group *B* and after /l/ in *Allah* [19].

Similarly, /AH:/, /An:/, /In:/, and /Un:/ are the long versions of /AH/, /An/, /In/, and /Un/ respectively.

The allophones /AY/, /AY1/, /AY2/, /OY/, /UW1/, /AW/, /AW1/ are actually two concatenated vowels in which the articulators move from one position to another. These vowels are called Diphthongs and they do not have corresponding English or Arabic phonemes. The allophones /AY2/ and /AY1/ appear when the phoneme /a/ comes before an unvoweled *Ya'a* ى . Similarly, /AW/ and /AW1/ appear when the phoneme /a/ precedes an unvoweled *Waw* و . The allophone /OW/ can be used for the vowel /u/ after a consonant of group *A* or *B*. Finally, the case of pharyngealized /i:/ is represented by the allophone /IX/.

2.4. Voiced Stops

/b/ has 6 allophones. /BB1/ and /BB2/, the normal unvoweled and voweled versions, are similar to their Engligh counter parts. However, in Arabic /BB1/ appears usually before voiceless fricatives. /BB3/, a clearly aspirated version *Muqalqal,* comes as word final or as the closing sound of syllables. /B1/, /B2/, and /B3/ are the pharyngealized versions of the previous three allophones, occurring if there is a consonant of groups *A* or *B* in the same syllable. The allophones of /d/ are almost the same as their counter parts in English, however, here too a new allophone is introduced which correspond to the unvoweled *Muqalqal* version of the /d/. Finally, *Dhad* /D/ is represented by three allophones for voweled cases and two allophones for the unvoweled versions.

2.5. Voiceless Stops

The Arabic voiceless stops as /t/ and /k/ are basically similar to the English counterparts. However, additional allophones are introduced to account for the coarticulations caused especially by the back consonants. /TT4/ is similar to /TT1/ but it is more aspirated. /TT3/ is similar to /TT2/ but it is more pharyngealized. On the other hand, *Ta'a* /T/ is represented by an unvoweled and three voweled allophones.

Similarly, *Qaf* /q/ is represented in the table by five allophones. For example, /q/ in *Qeel* قيل is acoustically different from /q/ in *Qoomo* since the vowel in the first case is articulated in the front of the mouth and in the latter is articulated in the back. The first three allophones of /k/ are quite similar to their English versions. /KK4/, on the other hand, is introduced to represent the unvoweled version after a letter of groups *A* or *B*. /KK5/ appears as a starting sound of syllables when it is followed by a letter of group *A* or *B*. /EE1/ and /EE2/ are closing *Hamzah* /E/ sound with /EE2/ after a pharyngealized vowel. /EE3/, /EE4/, and /EE5/ represent initial cases of /E/.

2.6. Voiced Fricatives

The Arabic voiced fricatives have basically one acoustic version and are relatively easy to isolate. However, the table contains at least two versions for each which differ mainly in their duration. For example, /AN1/ may be used as closing sound of a syllable or the first sound in gemination, while /AN2/ may then be used for the second voweled part of the gemination. An exception to this is /AN3/ which sometimes occurs if there is an emphatic consonant in the same syllable.

2.7. Voiceless Fricatives

The voiceless fricatives allophones of the phonemes /f/, /Θ/, /s/, /γ/, and /h/ are identical to their English counterparts. The longer version can be obtained by doubling the same phoneme. The allophone /SH1/ is used if the phoneme /ʃ/ comes unvoweled in the same syllable with a consonant of groups *A* or *B*. The *Sadh* /S/ allophones, /S1/ and /S2/, are almost acoustically the same but differ only in the duration. The two allophones of /x/ also differ in the same way. /H/ may be used as the closing sound of syllables or first sound in geminations. /H1/ may be used as the closing sound of syllables containing consonants of groups *A* or *B*. /H2/ is the regular voweled allophone. /H3/ is acoustically different because of the lip rounding in anticipation of the vowel /u/.

2.8. Affricatives

The only Arabic affricative is the *Geem* /dz/ which has three acoustically different allophones, /JH2/ occurs in word final, /JH1/ in gemination, while /JH/ is the normally used allophone for the voweled letter.

2.9. Nasals

Nasals are produced by completely blocking the oral cavity and allowing the air to pass through the nasal cavity. The table of the nasal allophones contains several new members; /MM:/ which is a version of /m/ that appears especially in gemination. /NM/ is a nasal sound between /n/ and /m/. /M/ and /M:/ are short and long pharyngealized versions of /m/ /n/ is characterized by its relatively large numbers of allophones. /NN1/ and /N1/ is the normal allophone and its pharyngealized version respectively. /NN2/ and /NN3/ appear before the pharyngealized /u/ vowel and the regular /u/ vowel (short or long) respectively. /NN:/ is the long version of /NN1/ occurs if an unvoweled /n/ comes before group E. /NG/ when an unvoweled /n/ comes before /k/, /q/, or /dz/. Finally, /NN4/ before group I. Longer allophones as /NN:/ and /N1:/ may be obtained by doubling their short versions.

2.10. R-Colored

The R-colored table of allophones reflects the coarticulation effect of /r/ on its preceding vowels. In Arabic, the R-coloring occurs only when an unvoweled *Ra'a* /r/ comes after a voweled consonant. Accordingly, the new members in this table account for the different combinations of the Arabic vowels and /r/.

2.11. Resonants

Resonants (semivowels) have similar acoustic characteristics to vowels because they are formed on an open tract, like vowels, and they are also highly resonant, like them. The Arabic resonants are basically the same as in Engligh except for a few new allophonic versions. /WW:/ appears when a geminated *Waw* /w/ comes after an unvoweled *Waw*. /WW2/ is a pharyngealized version which occurs when /w/ exists with an emphatic letter in the same syllable. /rr3/ is the emphatic version of /r/. /LL1/ is an emphatic version of /l/ as in *Allah*.

Finally, /YY3/ is the pharyngealized allophone of /y/.

2.12. Pharyngeal Controlled

Similar to the above R-colored allophones, these speech segments reflect the coarticulation effect of the emphatic and pharyngeal consonants of groups *A* and *B* on its preceding vowels. Each of these segments is basically a *VC* cluster where *C* is an emphatic or pharyngeal consonant of groups *A* and *B*, and *V* is a pharyngealized short vowel. For each consonant, four segments are presented; one with pharyngealized /a/, another with a pharyngealized /i/, a third with pharyngealized /u/, and the last one comes with a semi pharyngealized /a/ which normally appears after a non-pharyngealized consonant or /a/ short vowel.

The pharyngealization of the consonants /x/, /L/, and /ɣ/ is less pronounced when they are in the neighborhood of /u/ and /i/ and can be ignored with minor perceptual effect, especially in classical Arabic language. Hence, the pharyngeal controlled segments for these three consonants are presented with the vowel /a/ only.

3. PRONUNCIATION RULES

The pronunciation of the classical Arabic language is very structured and follows well established rules. In fact, this unique characteristic enables millions of non-Arabic speaking people to read Qur'an fluently, and in many cases even better than native Arabs, whose speech could actually be influenced by their dialect.

As mentioned in the introduction, there are several levels of rules that need to be developed and implemented for a successful text-to-speech system. These rules include: lexicon lookup; breathing rules; prosodic rules; pronunciation rules; and the synthesis model rules.

In this study we will consider only the pronunciation rules. These rules are responsible for converting a fully diacritized text to a string of allophones. The rules given here are purely phonetic rules which do not require morphological information, for example the rules related to pronunciation of the initial *Hamza* are not included since they depend on whether the word is noun or verb, and on the form of the verb.

The pronunciation rules are divided into two subgroups;

1. *Letter to sound rules:* which convert the text into an abstract ASCII string, called a phonemic

transcript, corresponding to the way the text is actually pronounced. This group includes also rules for parsing the text into syllable patterns.

2. *Allophone selection rules:* which resolve further the cases where the sound distinction can not be effected by the phonemic description alone; *e.g.* the short vowel sound /a/ when it comes after the /b/ consonant sounds dramatically different when it comes, emphasized, after an emphatic consonant as /T/ ط .

A computer program which implements some of the following pronunciation rules, called *Salpha* (Software for Arabic Language PHonetic Analysis), was presented in [4]. The following set of rules governs consonant/vowel combinations, usage of short and long vowels, effect of emphasizing and pharyngealization, pronunciation of /l/ in the particularization, rules for word final aspiration, *Tanween* and *Ghonna* rules, rules for conversion, suppression and assimilation, and rules for combining words. The presented set of rules can generate the allophone string correctly of about 98% of any fully diacritized Arabic text.

Before stating the rules, we will introduce first some notations to describe these rules. The notations were proposed in [26] to describe the English pronunciation rules. The notations were further refined in this study and adapted for the wider scope of the Arabic language framework.

The symbolic language presented here provided a shorthand for otherwise lengthy descriptive statements; it also helped in discovering many exceptions, and ambiguities, provided a systematic procedure for classifying the rules, and revealed the rules of interaction and nesting properties.

Each rule has the form:

$$D \leftarrow C[B]A \ .$$

Using the Arabic right-to-left convention, the meaning is that the character string B, occurring with right context A and left context C becomes D. A rule may also be specified in the form

$$D \leftarrow C[x]A \ , \qquad x \in X$$

which is now read "for any character string x of the set X, occurring with right context A and left context C becomes D".

Some special symbols are also defined to denote classes of strings or punctuations. A list of these symbols is shown in the table below.

For example, a typical rule is

$$@\text{هـ} \leftarrow @\,\underline{\text{d}}\ [\text{ـهـ}]$$

#, X	consonant, a consonant element of a specific set.
##	two or more consonants.
\$, Y	vowel, a vowel element of a specific set.
\$\$	two or more short vowels, or a long vowel.
@	A stop at an end of a single breath utterance.
^	A pause between words.
~ , –	End of a syllable.
( )	A geminated consonant.
=	same as ←.
d	any diacritics.
n	nunation (a suffix of an, in, or un).
ـَ	*Fat'ha* (short vowel /a/).
ـِ	*Kasra* (short vowel /i/).
ـُ	*Damma* (short vowel /u/).

Now, let $<S>$ be the set of all Arabic characters and diacritics, and let $<P>$ be the set of all Arabic language allophones. Define $\{S\}$ as the set of all ordered finite strings whose elements are in $<S>$, and define $\{P\}$ as the set of all ordered strings whose elements are in $<P>$.

Two elementary operations are defined on a string S, insertion of a consonant/diacritic and deletion of a consonant/diacritic. A Text Rule (TR), is a letter-to-sound rule consisting of a finite number of elementary operations such that:

1. for every $S1 \in \{S\}$, there exists $S2 \in \{S\}$.
2. $TR[S1] = S2$ (the Rule TR operates on the string $S1$, and produces $S2$).
3. $TR[S2] = S2$.

A mapping from $\{S\}$ to the set $\{P\}$ is called an Allophone Selection Rule (AR), if $AR[S] = P$.

Two text rules, $R1$ and $R2$, are said to be Independent if

$$R1[R2[S1]] = R2[R1[S1]].$$

Thus if two rules are independent, the result is invariant to the order of their applications.

A text rule $R2$ is said to Precede $R1$ if:

$$R1[R2[S]] \neq R2[R1[S]]$$

and

$$R2[R1[R2[S]]] = R1[R2[S]].$$

This relation implies that $R2$ must operate first on the text.

3.1. Letter-to-Sound Rules

The rules are listed below according to their preceding order. Independent rules are listed in arbitrary order.

1. Read one utterance of the text with diacritics. We assume full diacritics are available. Whenever *Hamza* /E/ is to be pronounced it is clearly indicated. Elidable *Hamza* and *Alif* are written with no diacritics, *e.g. Baab, Fasber,*

بَـابْ

فَـاصْبِـرْ

2. If a consonant is modified by a *Fat'ha,* the sign of /a/ vowel, and the next letter is unvoweled *Ya'a* ى /y/ in a word-final position, the /y/ letter is converted to an *Alif, e.g.*

عَـلَى ← عَـلاَ

ـَ [ى]^ ← # ـَـ ا ^

3. If an unvoweled *Waw* /w/ comes before *Alif* in a word final position, *Alif* shall be deleted, *e.g.*

ذَهَبُـوا ← ذَهَبُـو

.. ـُـ و [ا]^ ← . . . ـُـ و ^

4. If a long vowel comes before the definite article then the long vowel is reduced to a short vowel, *e.g.*

فِي المدينة ← فِـ المدينة ـِـ [ى]^ الـ . . . ← ـِـ ^ الـ

ذُو الحِجه ← ذُ الحِجه ـُـ [و]^ الـ ← ـُـ ^ ال . . .

5. If the last consonant of a sentence or utterance is *Ta'a* /t/ in the form ـة , it is converted to /h/ هـ regardless of its diacritic, *e.g.*

@ هـ . . . ← @ d [ـة]

بركـةُ ← بركه

6. The last consonant of a sentence or utterance is pronounced unvoweled. This rule applies regardless of the diacritic of the letter except if the last letter's diacritic is the suffix "an". In this case the stop should be with a normal long *Fat'ha* vowel, *e.g.*

ضرباً ← ضربا # [ـً] @ ← # ـَـ ا @

جميلٌ ← جَميلْ # [d] @ ← @ #

7. If the definite article comes at the beginning of a sentence or a new utterance, the *Hamza* "ء" shall be pronounced with *Fat'ha.* However, if the definite article comes after another word or conjunction (*Wasl*) the *Hamza* will be elided, *e.g.*

@ [أل] # . . . ← ء ـَـ ل # . . .

^ وَ [أل] # . . . ← ^ و ـَـ ل # . . .

8. If the definite article precedes a consonant which is pronounced roughly in the same articulation area as *Lam* /l/ or behind the upper teeth (the *Shamsi* letters, group *F*), the letter *Lam* is elided, and the next letter is geminated, *e.g.*

الـشَـمْـسُ ← أَشَّـمْـسُ

$x \in F$ ^ [ال] x ← . . . ـَـ (x) . . .

9. If the definite article precedes a consonant which is not *Shamsi* (called *Qamari,* group *G*), the letter *Lam* will be unvoweled, *e.g.*

وَالـقَمر ← وَلْـقَمر

$x \in G$ ^ [ال] x ← . . . ل x . . .

10. The *Tanween* diacritics (indicating a "un", "in", or "an" suffix) should be replaced by explicit suffixes, *i.e.* a short vowel sign and an unvoweled /n/, *e.g.*

. . . # [ـً] ^ ← . . . # ـَـ ن ^

. . . # [ـٌ] ^ ← . . . # ـُـ ن ^

11. If the letter *Ba'a* /b/ comes after an unvoweled *Noon* /n/, the letter /n/ is usually converted to the letter *Meem* /m/.

عَـنْـبَر ← عَـمْـبَر

. . . [ن] ب $ ← . . . م ب $

12. If an unvoweled /n/ precedes a voweled *Ra'a* /r/, then the /n/ sound is elided and /r/ is geminated, *e.g.*

مَن رَأى ← مرَّأى

[ن] ^ ر $ ← (ر) $ . . .

13. If an unvoweled /n/ precedes a voweled *Lam* /l/, then the /n/ letter is elided and *Lam* is geminated, *e.g.*

مَن له ← مَلَّه

[ن] ^ ل $ ← . . . (ل) $ . . .

14. If an unvoweled /D/ precedes *Tä'a* /T/, /D/ is converted to unvoweled /T/, *e.g.*

اضْطَرَّ ← اطَّرَّ

[ض] ط $ ← (ط) $ . . .

15. If an unvoweled /D/ (/d/), precedes a voweled /t/, the /D/ (/d/) and /t/ are replaced by a geminated /T/ (/t/), *e.g.*

أفَضْتُم ← أفَطُّم . . . [ض ت] $ ← (ط) $ . . .

أرَدتُم ← أرَتُّم . . . [د] ت $ ← (ت) $ . . .

16. If a middle word starts with an *Alif* (without *Hamza* and not part of the definite article) this *Alif* is elided. The next consonant will be unvoweled, *e.g.*

وَارْحمنا ← وَرْحمنا

. . . # $ # . . . ← . . . # [ا] ^ $ # . . .

17. If a letter is geminated, it is split into two consonants; the first one is unvoweled, while the second consonant is voweled according to the sign of the gemination (*Tashdeed*), *e.g.*

فـلاَّح ← فلْ لاَح

$ x ~ x ← . . . $ (x)

18. The consonant /L/ appears in the glorious name of *Allah*. If the glorious name of *Allah* is preceded by the vowel /i/, then the regular consonant /l/ should be used in the glorious name.

19. No more than two consonants can occur together, whether within a single word or at the end of one word and the beginning of the next word. If such a situation occurs (three unvoweled consonants in sequence) a "helping vowel" is introduced between the first consonant and the second consonant. If the last vowel before the first consonant is *Fat'ha*, the helping vowel is short *Kas'ra*. Conversely, if the last vowel before the first consonant is *Kasra*, the helping vowel is *Fat'ha*. Similarly, if the last vowel before the first consonant is *Damma*, the helping vowel is *Damma*, *e.g.*

مِنْ القمر ← مِنَ القمر ـِ [###] ... ← ـِ # ـَ ## ...

مَنْ القائد ← مَنِ القائد ـَ [###] ... ← ـَ # ـِ ## ...

أنتُم القادة ← أنتمُ القادة ـُ [###] ... ← ـُ # ـُ ## ...

20. Arabic text breaks down into only five syllable patterns; CV, CVV, CVVC, CVC, and CVCC. Each voweled consonant occurs at the beginning of a new syllable. Each geminated letter occurs at the end of a syllable and starts a new syllable. Each syllable contains only one short or one long vowel.

(In Qur'anic recitation there are a number of prolonged syllables, *e.g.* CVVVV, CVVVVC, and CVVVVVVC. The CVVVV replaces CVV if it precedes a *Hamza*. The CVVVVVVC replaces CVVC in the middle of words or sentences, and CVVVVC may replace CVVC at the end of an utterance. Other rules for prolonging the CVV and CVVC syllables can be deduced from [24]).

21. If /h/ in the form ـه comes in a word final position in a CV syllable and before a CV or CVV syllable, the short vowel of /h/ is converted to a long vowel by insertion of a proper long vowel letter such as *Alif*, *Waw*, or *Ya'a*, *e.g.*

إنَّهُ هُوَ ← إنَّ هُوْ هُو

~ $ # ^ Y [ـه] ← ~ $ # ^ YY هـ . . .

22. If *Hamza* /E/ comes as the initial and final consonant in a CVC or CVVC syllable, the syllable is replaced by CVV syllable, *i.e.* the closing *Hamza* will be elided, *e.g.*

لَـمْ أأكُل ← لَـمْ آكُل

~ YY ء . . . ← . . . ~ [ء] Y ء ~

3.2. Allophone Selection Rules

Next we apply rules which further convert the abstract transcription into allophone strings. The rules are again arranged in a precedence order. Independent rules, as in the text rules, may be rearranged. The allophone rules basically reflect the coarticulation effects. For example, semivowels/vowels interaction, pharyngealization, stops/stops interaction, emphatic/vowels and vowel/emphatic interaction, the anticipatory effect of /u/ which causes the consonant preceding it to be pronounced with rounded lips; change of /n/ by group *E* and *I*; R-controlled vowels, and nasalization of vowels.

The simplest way to implement these rules is by direct mapping look up tables. The dominant factor in consonant selection is its syllabic position as the first, second, or third consonant in a syllable. The allophones of a consonant are categorized here into over 22 possible variants as follows:

(C1) Before the /a/ vowel.
(C2) Before the /i/ vowel.
(C3) Before the /u/ vowel.
(C4) Unvoweled short, as in CVCC syllable.
(C5) Unvoweled devoiced in a closing sound of a syllable before group $\{J, K\}$.
(C6) Unvoweled long in a closing sound of a syllable NOT before group $\{J, K\}$.
(C7) Emphasized (pharyngealized) before an emphasized /a/ vowel.
(C8) Emphasized (pharyngealized) before an emphasized /i/ vowel.
(C9) Emphasized (pharyngealized) before an emphasized /u/ vowel.
(C10) Emphasized (pharyngealized) unvoweled short as in CVCC syllables.
(C11) Emphasized (pharyngealized) unvoweled in a closing sound of a syllable before group $\{J, K\}$.
(C12) Emphasized (pharyngealized) long in a closing sound of a syllable and NOT before group $\{J, K\}$.
(C13) In a VC cluster with a short /a/ vowel after a non emphatic consonant.
(C14) In a VC cluster with a short /i/ vowel after a non emphatic consonant.
(C15) In a VC cluster with a short /u/ vowel after a non emphatic consonant.
(C16) In a VC cluster with a short /a/ vowel after a consonant of group *A*.
(C17) In a VC cluster with a short /a/ vowel after a consonant of group *B*.
(C18) In a VC cluster with a short /i/ vowel after an emphatic consonant.
(C19) In a VC cluster with a short /u/ vowel after an emphatic consonant.
(C20) Special cases.
(C20) Special cases.

A table is constructed wherein the allophone entries are selected from the available consonant's allophones, as depicted in Table 1. For example, the table entries for the phoneme /b/ will be:

C1	C2	C3	C4	C5	C6
/BB2/	/BB2/	/BB2/	/BB1/	/BB1/	/BB3/

C7	C8	C9	C10	C11	C12
/B2/	/B2/	/B2/	/B1/	/B1/	/B3/

C13	C14	C15
/AE/ /BB1/	/IH/ /BB1/	/UH/ /BB1/

C16	C17	C18	C19
/AA/ /B1/	/AH/ /B1/	/IX/ /B1/	/UX/ /B1/

C20	C21
----	----

Similarly we construct a table for vowels:

(V1) /a/ short regular /AE/.
(V2) /i/ short regular /IH/.
(V3) /u/ short regular /UH/.
(V4) /a/ short transient /AA1/.
(V5) /a/ short emphasized /AH/ after group *A*.
(V6) /a/ short emphasized /AA/ after group *B*.
(V7) /i/ short emphasized /IX/.
(V8) /u/ short emphasized /UX/.
(V9) /a/ short nasalized /An/.
(V10) /i/ short nasalized /In/.
(V11) /u/ short nasalized /Un/.
(V12) /a/ long regular /AE:/.
(V13) /i/ long regular /IY/.
(V14) /u/ long regular /UW2/.
(V15) /a/ long emphasized /AH:/ after group *A*.
(V16) /a/ long emphasized /AA:/ after group *B*.
(V17) /i/ long emphasized /IX:/.
(V18) /u/ long emphasized /UX:/.
(V19) /a/ long nasalized /An:/.
(V20) /i/ long nasalized /In:/.
(V21) /u/ long nasalized /Un:/.

3.2.1. Consonant Rules for Classical Arabic

1. In the CV, CVVC, or CVV syllables, the initial consonant of the syllables is replaced by the allophone C1, C2, or C3, depending on whether the following vowel is /a/, /i/, or /u/, respectively.

هَامَ ← [هـ] ـَـ ـَـ – م ← /HH1/

هُنا ← [هـ] ـُـ – ن ـَـ ← /HH2/

2. In the CVC or CVCC syllables, if any of the closing consonants is emphatic of group A, then initial consonant of the syllable is replaced by the allophone C7, C8, or C9 depending on whether the following vowel is /a/, /i/, or /u/, respectively. In all other cases, the initial consonant of the syllables is replaced by the allophone C1, C2, or C3, depending on whether the following vowel is /a/, /i/, or /u/, respectively.

مَنزِل ← [م] ـَـ ن – ز ـِـ ل ← /MM1/

مَضْرَب ← [م] ـَـ ض – ر ـَـ ب ← /M1/

3. In a CVC, CVVC, and CVCC syllables, if the initial consonant is of group $\{A, B\}$, the VC cluster is replaced by C17, C18, or C19 depending on whether the preceding vowel is /a/, /i/, or /u/, respectively.

ضَيف ← ض [ـَـ ى] ف ← /AY2/

صَقْر ← ص [ـَـ ق] ر ← /AQ/

4. In the CVC, CVVC, and CVCC syllables, if the initial consonant is NOT of group $\{A, B\}$, the VC cluster is replaced by C13, C14, or C15 depending on whether the preceding vowel is /a/, /i/, or /u/, respectively.

بَاب ← ب ـَـ [ـَـ ب] ← /AE/ /BB1/

سَيف ← س [ـَـ ى] ف ← /AY1/

5. The closing consonant of a syllable is replaced by C6. However, if there is one or more emphatic consonant in the same syllable, it is replaced by C12.

كَلْبْ ← ك ـَـ ل [ب] ← [BB3]

ضَرب ← ض ـَـ ر [ب] ← [B3]

3.2.2. Consonant Rules for Modern Standard Arabic

1. In a CV or CVV syllable comes after a consonant of group A, the initial consonant of the syllable is replaced by the allophone C7, C8, or C9 depending on whether the following vowel is /a/, /i/, or /u/, respectively. Otherwise, the initial consonant of the syllables is replaced by the allophone C1, C2, or C3, depending on whether the following vowel is /a/, /i/, or /u/, respectively.

ضَبابُ ← ض ـَـ – [ب] ـَـ ـَـ – ب ـُـ ← /B2/

كبابُ ← ك ـَـ – [ب] ـَـ ـَـ – ب ـُـ ← /BB2/

2. If a CVC, CVVC, or CVCC syllable comes after an unvoweled consonant of group A, or at least one of its closing consonants is an emphatic from groups $\{A, B\}$, then the initial consonant of the syllable is replaced by the allophone C7, C8, or C9 depending on whether the following vowel is /a/, /i/, or /u/, respectively. In all other cases, the initial consonant of these syllables is replaced by the allophone C1, C2, or C3, depending on whether the following vowel is /a/, /i/, or /u/, respectively.

بَطَّة ← [ب] ـَـ ط – ط ـَـ ه ← /B2/

بَرُّ ← [ب] ـَـ ر – ر ـُـ ← /B2/

3. In a CVC, CVVC, and CVCC syllables, if the initial consonant is of group $\{A, B\}$ or emphasized, the VC cluster is replaced by C16, C18, or C19 depending on whether the preceding vowel is /a/, /i/, or /u/, respectively. Otherwise, the VC cluster is replaced by C13, C14, or C15 depending on whether the preceding vowel is /a/, /i/, or /u/, respectively.

ضَبابْ ← ض ـَـ – ب ـَـ [ـَـ ب] ← /AA/ /B1/

سَبتْ ← س [ـَـ ب] ت ← /AE/ /BB1/

4. If the closing consonant of a syllable does not precede a consonant of group $\{J, K\}$, it is replaced by C6. However, if there is one or more emphatic consonant in the same syllable, it is replaced by C12.

حِزب ← ح ـِـ ز [ب] ← /BB3/

ضباب ← ض ـَـ – ب ـَـ [ـَـ ب]

$$/AA/\ /B3/ \xleftarrow{\text{RULE}\ 3} /AA/\ /B1/$$

5. If the closing consonant of a syllable precedes a consonant of group $\{J, K\}$, it is replaced by C5.

However, if there is one or more emphatic consonant in the same syllable, it is replaced by C11.

سَحْب ← س ـَـ ح [ب] ← /BB1/

ثُقب ← ث ـُـ ق [ب] ← /B1/

3.2.3. Some Vowel Rules

1. In the CV (CVV) syllables, if the initial consonant is of group *A*, then the vowel is represented by the allophones V5, V7, or V8, respectively (V15, V17, or V18, respectively) depending on whether the vowel is /a/, /i/, or /u/, respectively.

ضالّ ← ض [ـَـ] − ل . . . ← /AH:/

قَليل ← ق [ـَ] − ل . . . ← /AH/

2. In the CV (CVV) syllables, if the initial consonant is of group *B* or positionally emphasized, then the vowel is represented by the allophones V6, V7, or V8, respectively (V16, V17, or V18, respectively) depending on whether the vowel is /a/, /i/, or /u/, respectively.

رَبى ← ر [ـَـ] − ب . . . ← /AA/

غابَ ← غ [ـَـ] − ب ـَـ ← /AA:/

3. In the CV (CVV) syllables, if the initial consonant is NOT of group $\{A, B\}$ and is NOT positionally emphasized, then the vowel is represented by the allophones V1, V2, or V3, respectively (V12, V13, or V14, respectively) depending on whether the vowel is /a/, /i/, or /u/, respectively.

سَمِعَ ← س [ـَـ] − م ـِـ − ع ـَـ ← /AE/

سَاعَدَ ← س [ـَـ] − ع ـَـ − . . . ← /AE:/

4. In the CVVC syllables, if the initial consonant is of group *A*, then the long vowel is represented by two short allophones, the first one is represented by V5, V7, or V8 respectively depending on whether the vowel is /a/, /i/, or /u/, respectively. The second vowel is determined by the VC cluster of the closing consonant.

ضَار ← ض [ـَـ] ـَـ ر ← /AH/

قِيلَ ← ق [ـِـ] ـِـ ل ← /IX/

5. In the CVVC syllables, if the initial consonant is of group *B* or positionally emphasized, then the long vowel is represented by two short allophones V6, V7, or V8, respectively depending on whether the vowel is /a/, /i/, or /u/, respectively. The second vowel is determined by the VC cluster of the closing consonant.

رَاح ← ر [ـَـ] ـَـ ح ← /AA/

غُول ← غ [ـُـ] ـُـ ل ← /UX/

6. In the CVVC syllables, if the initial consonant is NOT of group $\{A, B\}$, and is NOT positionally emphasized, the vowel is represented by two short allophones; the first one is the allophones V1, V2, or V3, respectively depending on whether the vowel is /a/, /i/, or /u/, respectively. The second is determined by the VC cluster of the closing consonant.

سام ← س [ـَـ] ـَـ م ← /AE/

تُوت ← ت [ـُـ] ـُـ ت ← /UH/

7. In the CVCC syllables, the vowel is determined by the VC cluster as discussed before.

سَيف ← س [ـَـ ى] ف ← /AY1/

صَقر ← ص [ـَـ ق] ر ← /AQ/

3.2.4. Exceptions

1. If a guttural letter of group *E* comes after an unvoweled *Noon* /n/, the allophone /NN1/ should be used for the /n/ sound.

من أنا ← م ـَـ [ن] ^ ء ـَـ . . . ← /NN1/

أنهار ← ء ـَـ [ن] هـ ـَـ . . . ← /NN1/

2. If an unvoweled *Noon* /n/, comes before a letter of group *I*, the allophone /NN3/ should be used for the /n/ sound.

إنتهى ← ء ـِـ [ن] − ت . . . ← /NN3/

ينشُر ← ى ـَـ [ن] − ش . . . ← /NN3/

3. If an unvoweled *Noon* /n/, comes before a letter of group *L*, the allophone /NG/ should be used for the /n/ sound.

ينكر ← ى ـ [ن] – ك . . . ← /NG/

ينجَرح ← ى ـ [ن] – ج . . . ← /NG/

4. If an unvoweled *Noon* /n/, comes before a *Ya'a* /y/, the allophone /NY/ should be used for the /n/ sound.

مَن يعمل ← م ـ [ن] ^ ى . . . ← /NY/

5. If an unvoweled *Noon* /n/, comes before letter /b/ or /m/, the allophone /NM/ should be used for the /n/ sound.

مِن ما ← م ـ [ن] ^ م . . . ← /NM/

مَن بات ← م ـ [ن] ^ ب . . . ← /NM/

EXAMPLE

بِسمِ اللهِ الرَحمَنِ الرَحِيم

إقرَأ بِسمِ رَبِّكَ الَّذِى خَلَق .

خَلَقَ الإنْسَانَ مِنْ عَلَق .

إقْرَأ وَرَبُّكَ الأكْرَم .

الَّذِى عَلَّمَ بِالقَلَم .

عَلَّمَ الإنْسَانَ مَالَم يَعلَم .

ب ـ س – م ـ ل – ل ـ ـ ـ – هـ ـ ـ ر – ر ـ ح – م ـ ـ ـ –

ن ـ ـ ر – ر ـ – ح ـ ـ ـ م @

ء ـ ـ ق ر ـ ء ^ ب ـ ـ س – م ـ ـ ^ ر ـ ب – ب ـ ـ – ك ـ ـ ل –

ل ـ ـ – ذ ـ ـ ـ ^ خ ـ ـ – ل ـ ـ ق @

خ ـ ـ – ل ـ ـ – ق ـ ـ ل – ء ـ ـ ن – س ـ ـ ـ – ن ـ ـ ^ م ـ ـ ن ^

ع ـ ـ – ل ـ ـ ق @

ء ـ ـ ق – ر ـ ـ ء ^ و ـ ـ ^ ر ـ ـ – ب – ب ـ ـ – ك ـ ـ ل –

ء ـ ـ ك – ر ـ ـ م @

ء ـ ـ ل – ل ـ ـ – ذ ـ ـ ـ ^ ع ـ ـ ل – ل ـ ـ – م ـ ـ ^ ب ـ ـ ل –

ق ـ ـ – ل ـ ـ م @

ع ـ ـ ل – ل ـ ـ – م ـ ـ ل – ء ـ ـ ن –

س ـ ـ ـ – ن ـ ـ ^ م ـ ـ ^ – ل ـ ـ م ^ ي ـ ـ ع – ل ـ ـ م @

/BB2/ /EH/ /SS/ /PA2/ /MM/ /IH/ /LL/ /PA1/ /LL/ /AE:/ /PA2/ /HH1/ /IR1/ /PA1/ /RR3/ /AA/ /H3/ /PA2/ /MM/ /AE:/ /PA1/ /NN1/ /IR1/ /PA1/ /RR3/ /AA/ /PA2/ /H1/ /IH:/ /MM/ /PA5/.

/EE4/ /IQ/ /PA2/ /RR3/ /AA/ /EE2/ /PA3/ /BB2/ /EH/ /SS/ /PA2/ /MM/ /IH/ /PA3/ /RR3/ /AA/ /B3/ /PA1/ /BB2/ /EH/ /PA2/ /KK1/ /AE/ /LL/ /PA1/ /LL/ /AE/ /PA2/ /ZH3/ /IY/ /PA3/ /KH/ /AA/ /PA2/ /LL/ /aQ/ /PA5/ /PA5/.

/KH/ /AA/ /PA2/ /LL/ /AE/ /PA1/ /QQ2/ /AH/ /LL1/ /PA3/ /EE4/ /IH/ /NN1/ /PA2/ /SS/ /AE:/ /PA2/ /NN1/ /AE/ /PA3/ /MM/ /In/ /NN1/ /PA4/ /AN1/ /AE/ /PA2/ /LL1/ /aQ/ /PA5/.

/EE4/ /IQ/ /PA2/ /RR3/ /AA/ /EE2/ /PA3/ /WW/ /AE/ /PA3/ /RR3/ /AA/ /B3/ /PA1/ /BB2/ /UH/ /PA2/ /KK1/ /AE/ /LL/ /PA1/ /EE3/ /AE/ /KK2/ /PA3/ /RR3/ /AA/ /M1/ /PA5/.

/EE3/ /AE/ /LL/ /PA1/ /LL/ /AE/ /PA2/ /ZH3/ /IY/ /PA3/ /AN2/ /AE/ /LL/ /PA1/ /LL/ /AE/ /PA2/ /MM/ /AE/ /PA3/ /BB2/ /EH/ /LL/ /PA2/ /QQ2/ /AH/ /PA2/ /LL/ /AE/ /MM/ /PA5/.

/AN2/ /AE/ /LL/ /PA1/ /LL/ /AE/ /PA2/ /MM/ /AE/ /LL/ /PA2/ /EE4/ /EH/ /NN4/ /PA2/ /SS/ /AE:/ /PA2/ /NN1/ /AE/ /PA3/ /MM/ /AE/ /PA3/ /LL/ /AE/ /MM/ /PA3/ /YY2/ /AE/ /AN1/ /PA2/ /LL/ /AE/ /MM/.

4. CONCLUSION

The paper describes a combined Arabic/English allophone set and a comprehensive set of Arabic pronunciation rules based on a proposed allophone set. This work represents the basic skeleton of an effort to construct an allophone based Arabic text-to-speech system by rules. One of the important missing blocks is the implementation of the Arabic prosodic rules. However, much work is still needed on Arabic to deduce its prosodic rules. Another drawback of our experimental system is its LPC synthetic model. A formant based model is currently being developed with a much wider frequency spectrum to replace the LPC model. Another major difficulty is the need for a fully diacritized text. Unfortunately, most modern written Arabic, as in books and newspapers, has very limited diacritics. Automatic vowelization of unrestricted Arabic text requires a comprehensive morphology, syntax, and semantic analyzer; a challenging task for the 90s.

ACKNOWLEDGEMENT

The author would like to acknowledge the reviewers for their valuable comments, which contributed to the clarity of the paper, and in particular for their suggestions for the statements of rules 4, 6, 19, and 22.

REFERENCES

[1] M. Elshafei Ahmed, "A Device for the Programmed Teaching of Arabic Language and Recitations", *U.S. Patent No. 4 710 877,* 1987.

[2] M. Elshafei Ahmed, "Allophone-Based Arabic Speech Synthesis", *Proceedings of the 11th National Computer Conference, Dhahran,* 1989, p. 330.

[3] M. Elshafei Ahmed, M. S. Ahmed, and A. M. Marzooq, "An Arabic Text-to-Speech System", *Proceedings of the Regional Conference on Informatics and Arabization, Tunis,* March 1988.

[4] M. Elshafei Ahmed and A. Al-Sultan, "A Letter-to-Sound Algorithm for Automatic Conversion of Voweled Arabic Text to Speech", *Proceedings of the 11th National Computer Conference, Dhahran,* 1989, p. 401.

[5] M. F. Abu Alyazeed, M. M. R. Al-Ghoneimy, and M. F. Mohammad, "Comparison of the Syllable and Sub-Syllable Methods for Speech Synthesis", *Proceedings of the Second Conference on Arabic Computational Linguistics, Kuwait,* 1989, p. 424.

[6] Y. A. El-Imam, "Synthesis of Arabic Speech from Sub-Syllables", *Proceedings of the Regional Conference on Informatics and Arabization, Tunis,* March 1988.

[7] A. Mouradi, A. Rajouai, and M. Najim, "Unlimited Vocabulary Synthesis for Arabic Language", *Proceedings of the 4th Int. Conf. on Digital Signal Processing in Communications, Loughbourgh Univ., England,* April 1985.

[8] Y. A. El-Emam and A. Dannan, "The Phonetics of Modern Standard Arabic", *Technical Report # KSC 022, IBM Kuwait Scientific Center,* October 1986.

[9] Salem Ghazali, "Elements of Arabic Phonetics", *Proc. of the Arab Science and Technology on Applied Arabic Linguistics and Signal and Information Processing, Rabat, Morocco,* 1983.

[10] Salman H. Al-Ani, *Arabic Phonology — An Acoustical and Physiological Investigation.* Paris: Mouton, 1970.

[11] M. Maryati, "Speech Processing: Application to the Arabic Language", *Workshop on Computer Processing of the Arabic Language, Kuwait,* 1985.

[12] Suleyman S. Penbeci and Jasem A. Hejres, "Properties of Arabic Speech Signals for Human–Computer Communication", *Proceedings of the 9th National Computer Conference and Exhibition, Riyadh,* 1985.

[13] Abdullah Al-Dannan and B. Thalouth, "Comprehensive Arabic Morphological Analyser/Generator", *Technical Report # KSC 029, IBM Kuwait Scientific Center,* February 1987.

[14] A. El-Dessouki, A. Nazif, O. El-Dessouki, and M. Ahmed, "An Expert System for Understanding Arabic Sentences", *10th National Computer Conference & Exhibition, King Abdul-Aziz University, Jeddah,* 1987.

[15] T. A. Elsadany and M. A. Hashish, "Semi-Automatic Vowelization of Verbs", *10th National Computer Conference & Exhibition, King Abdul-Aziz University, Jeddah,* 1987.

[16] Y. Hilal, "Arabic Morphological Generator", *Proceedings of the Second Conference on Arabic Computational Linguistics, Kuwait,* 1989, p. 1.

[17] Mohammad Khayat, "Directions in Natural Arabic Understanding", *Proceedings of the Second Conference on Arabic Computational Linguistics, Kuwait,* 1989, p. 471.

[18] Saad K. Al-Jabri, "A Syntactic Analyzer for Arabic", *M.Sc. Thesis, College of Computer Science and Engineering, King Fahd University of Petroleum and Minerals,* June 1989.

[19] Kamal Ibraheem Badry, *Elm Allugha Almubarmag (Arabic Title).* Riyadh: King Saud University Press, 1983.

[20] Ibraheem Anees, *Alaswat Allaghawaiya (Arabic Title).* Cairo: Anglo Press, 1979.

[21] Khalid I. Semaan (Trans.), *Arabic Phonetic; Ibn Sina's Risala on the Point of Articulation of Speech Sounds (Sh. Muhammad Ashraf).* Lahore, Pakistan: xxxx, 1963.

[22] Ahmed Mukhtar Omar, *Derasat Alsawt Allaghawy (Arabic Title).* Cairo: Alam Alkutub, 1976.

[23] Nabil Ali, *Allugha Alarabia Walhasoob (Arabic Title).* xxxx: Alkhat Press, 1988.

[24] Abdulaziz Abdulfattah Alqari, *Altagweed Almuyassar.* Kuwait: Dar Alarqam Press, 1980.

[25] N. Umeda and C. H. Coker, "Allphonic Variations in American English", *J. Phonetics,* **2** (1976), p. 1.

[26] H. S. Elovitz, R. Johnson, A. McHugh, and J. E. Shore, "Letter-to-Sound Rules for Automatic Translation of English Text to Phonetics", *IEEE Transactions on Acoustics, Speech, and Signal Processing,* **ASSP-24(6)** (1976), p. 446.

[27] O. Fujimura, "Spectra of Nasalized Vowels", *Research Laboratory for Electronics, QPR 62, MIT, Cambridge,* 1960, p. 214.

[28] C. H. Coker, "A Model for Articulatory Dynamics and Control", *Proc. IEEE,* **64** (1976), p. 452.

[29] Dennis H. Klatt, "Review of Text-to-Speech Conversion for English", *J. Acoust. Soc. Am.,* **82(3)** (1987), p. 737.

[30] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals.* New York: Prentice Hall, 1978.

[31] J. D. Markel and A. H. Gray, *Linear Prediction of Speech.* Berlin: Springer–Verlag, 1976.

[32] D. J. Quarmby and J. N. Holmes, "Implementation of a Parallel Formant Speech Synthesizer Using a Single-Chip Programmable Signal Processor", *IEE Proceed.,* **131** (1984), p. 563.

Paper Received 29 May 1990; Revised 29 June 1991.