PERIODICA POLYTECHNICA SER. EL. ENG. VOL. 36, NO. 1, PP. 39–52 (1992)

ACOUSTIC BUILDING UNITS FOR FORMANT SYNTHESIS TEXT-TO-SPEECH CONVERTER SYSTEM FOR MODERN STANDARD ARABIC

M. A. SABAH,[1] G. GORDOS and G.OLASZY[2]

Department of Telecommunications and Telematics
Technical University, H-1521 Budapest

Received February 27, 1992.

Abstract

In this paper an inventory of acoustic building units (ABUs) used for the synthesis of Arabic speech is presented. The ABUs are generated for the free programmable PCF-8200 formant synthesizer chip which has been used in the development of the real time text-to-speech multilingual system, the MULTIVOX. To utilize these ABUs for the synthesis of Arabic speech a set of 36 Arabic sounds and all their possible combinations are defined. The inventory of 255 ABUs is designed that each sound combination can be built up by using some of those ABUs. A grapheme-to-phone-code converter is designed so to convert the written input text into its equivalent phone-codes. Furthermore, it contains solutions for the difficult phonetic problems in the Arabic input text.

Keywords: formant synthesis techniques, speech synthesis by rules, text-to-speech systems, phonetics of Arabic.

Introduction

Speech analysis and synthesis is an important topic in research today. This involves phonetics, speech acoustics computer engineering and signal processing scientists. These studies attempt to understand the vocal sounds and patterns inherent in languages so that speech may be recreated synthetically. Personal and mainframe computers equipped with general purpose signal processing unit allow researchers to segment, analyze and synthesize the speech for experiments before final commitment to a target synthesizer (EL-IMAM, 1987).

The most obvious way to output speech may seem to select the appropriate speech units and then record these units digitally. These units can then be concatenated to form the desired utterances. There are two type of such units, namely the long time duration speech units (e.g. words and sentences) and the short time duration speech units (e.g. syllables, diphones

[1] On leave from the Department of Electrical and Electronic Engineering, University of Assiut, Assiut, Egypt.
[2] Phonetics Laboratory

and phonemes). The use of the first type of speech units in the synthesis of unlimited vocabulary text requires huge storage, however, if the second type of units is used, it is difficult to deal with co-articulations and transitions between sounds. This has a very great influence on the naturalness of the produced speech. Therefore, for producing high-quality speech with minimum storage requirements, it is important to use short-time duration units and to be able to deal with the contents of these units to help in solving the co-articulation and transition problems. In the MULTIVOX system, microelement speech units (acoustic building units ABUs) have been used to solve both the storage and the co-articulation and the transition problems. The idea behind the microelements is that the acoustic realization of the human articulate motions can be represented in form of phonetic data (formant frequencies, formant bandwidths, voiced/unvoiced parameters, etc.). These phonetic data can be updated and interpolated every short-time (8-50 msec) interval to represent a short acoustic part of the speech waveform. Continuous speech can be generated by concatenated short acoustic parts (ABUs) and put them in digital form to represent the control codes of a synthesizer (OLASZY and GORDOS, 1989).

The Arabic language is spoken by more than 200 million people. It is quite evident that the ability to use the Arabic as a man–machine communication language seems to be a priority for all Arabs and may lead to a better development of management, industry and education in the Arab world. In the last few years, many scientific centers have developed the speech recognition and synthesis as well as the text-to-speech converter systems for Arabic language. The most famous centers in the Arab world are: IBM Cairo Scientific Center (SELIM and ANBAR, 1986), IBM Kuwait Scientific Center (EL-IMAM, 1989 & 1990) and Faculté des Sciences Morocco (MOURADI et al., 1985; RAJOUANI et al. 1987). However, as far as we know, real-time system is not available yet. The development of a real-time text-to-speech system for unlimited vocabulary of Arabic language is the subject of our paper.

In this article a set of 36 Arabic sounds are introduced for the purpose of developing a real-time text-to-speech converter system. A grapheme-to- phone-code converter is developed to convert the Arabic input text to suitable equivalent phone-codes. In the converter most of the pronunciation problems in the Arabic text are solved. Using a novel synthesis method based on the phoneme-to-phone-code segmentation and the replacement of each phone-code by a set of ABUs, Arabic speech with good quality is synthesized (Sabah et al. 1991) . The synthesis method was developed for the MULTIVOX text-to-speech converter system that is a patent product introduced in the Technical University of Budapest in co-operation with the Institute of Linguistics of the Hungarian Academy of Sciences. Based on our work, the MULTIVOX system that is originally assigned for European languages has been adopted for the synthesis of Modern Standard Arabic (OLASZY and GORDOS, 1987; OLASZY, 1989).

The system uses the general purpose free programmable PCF-8200 formant synthesizer chip for speech generation. Speech is formed by the concatenation of ABUs. Therefore an inventory of 255 ABUs is designed so that each phone-code-pair can be built up by concatenating some of those ABUs. Generally, the number of the ABUs required for the speech synthesis is language dependent. For Arabic language all of the 255 ABUs are used. The segmentation of the language by replacing each phone-code pair by a set of ABUs makes the synthesis method more flexible because it is possible to interchange and/or adapt any ABU in the ABUs sequence of the phone-code pair(s) to adjust the required utterance for any sound or sound combinations. By this method it is possible to deal with all the phonetic variations of the language and most of the phonemic co-articulations, too.

Arabic Letters, Sounds and Sound Combinations

The Arabic alphabet consists of 29 letters. From them 28 are consonants and only one is long vowel. The Arabic consonant sounds are distinguished by:

- voiceless glottal stop /e'/;
- voiced pharyngeal fricative /o:/ and voiceless pharyngeal fricative /h'/;
- voiced uvular fricative /g/ and voiceless unaspirated uvular stop /q/;
- pharyngealized consonants: /s'/, /d'/, /t'/ and /z"/ which are the voiceless post-dental sibilant emphatic consonant, the voiced post-dental emphatic stop, the voiceless post-dental emphatic stop and the voiced post-interdental emphatic fricative, respectively.

Although the Arabic language is characterized mainly by its consonant sounds (each represents by an alphabet), only one letter from the Arabic alphabets, the 'alif' represents the long vowel /ee/. It should be noted that, in some sequences two of the 28 consonants, namely the /w/ and /y/, serve also as long vowels (if the /w/ is preceded by the short vowel /o/, the /y/ is preceded by the short vowel /i/ and any of them is not followed by any short vowel) which are represented by the phonemes /oo/ and /ii/, respectively.

There are three short vowels /e/, /o/ and /i/ which contrast phonemically with their long counterparts /ee/, /oo/ and /ii/, respectively. It is noticed that the time duration of the long vowel is close to be twice of the corresponding short one. Arabic short vowels are not indicated by letters as in the case of other languages, but they are indicated by signs above or below the consonant carrying them. These signs are indicated as 'َ' , 'ُ' and 'ِ'. They are called 'elfatha', 'eldamma' and 'elkasra', and correspond phonemically to /e/, /o/ and /i/, respectively. Vowels, short and long, in Arabic are characterized by:

- Arabic vowels cannot be initials, but they may occur either between two consonants or at the end of the word. Here, it is important to be noted that those words which are normally considered to begin with a vowel, for example the Arabic name Ahmed, are usually initiated with a glottal stop before the vowel.
- Two different Arabic vowels cannot follow each other.
- In the process of Arabic speech, the phonetic quality of each vowel changes due to the effect of its phonetic environment, i.e. depends on the surrounding consonant(s).

Unfortunately, the vowel system of the written Arabic text is not compatible with the spoken one. The main reason of this difficulty is the effect of the Arabic consonants (specially the pharyngealized consonants) on the vowel sounds. Therefore, for the purpose of the present system, a set of 36 sounds, including short vowels only, are defined for Arabic speech. Among these sounds, one is devoted for the space at the beginning of each word, 28 are devoted for consonants, where each sound represents a phoneme and seven sounds represent the short vowels, namely 3 sounds are assigned to 'elfatha', 2 sounds to 'eldamma' and 2 sounds to 'elkasra'. Long vowels are considered to be of almost the double duration of that of the corresponding short vowels. The three short vowel sounds [e], [o] and [i], which are called non-pharyngealized vowels, are associated with the non-pharyngealized consonants. While the three short vowel sounds [a], [u], and [i'], pharyngealized vowels, are employed with the pharyngealized consonants {[s'], [d'], [t'], [z']}, the uvular stop [q] and the trill [r]. The seventh short vowel [ea] is utilized as a short vowel preceding the uvular [q] in some sequences. The [ea] sound is also used in the pronunciation of other words. The pharyngealized vowels and the vowel [ea] are not supported in the input text (they have no written symbols). They appear only in phone-code sequences. These sequences are manipulated by some rules established in the grapheme-to-phone-code converter that will be discussed in the next section.

In addition to the above features, Arabic speech has four diphthongs. They are formed when the glides /w/ and /y/ occur as nongeminate sounds after the short vowels [e] and [a]. These diphthongs are [ew], [aw], [ey] and [ay], respectively. Another feature of Arabic consonants is that they may be doubled in medial and final positions, more frequently in medial position. They are never doubled in initial position. This occurs by doubling the

Table 1-a
Arabic symbols, their corresponding phoneme symbols
(Keyboard symbols) and examples in both

Sound code	Symbols Key board	Symbols Ara bic	Examples in Keyboard symbols	Examples in Arabic symbols	Sound code	Symbols Key board	Symbols Ara bic	Examples in Keyboard symbols	Examples in Arabic symbols
1	-								
2	t'	ط	t'eee'ir	طائر					
			met'eer	مطار	10	e	ـَ	e'emel	أمل
			met"eet'	مطاط				meee'	ماء
2 2	t"	طّ	qit"eh	قطّه				ketebe	كتب
			xet"	خطّ	11	b	ب	bileed	بلاد
3	e	ـَ	e'et'feel	أطفال				jobn	جبن
			qelem	قلم				beeb	باب
			merd'ee	مرضى	12	z'	ذ	z'eheb	ذهب
4	o	ـُ	e'oh'ibb	أحبّ				heez'ee	هذا
			doleeb	دلاب				monz'	منذ
			yektobo	يكتب	12 12	zz'	ذّ	o:ezz'ebe	عذّب
5	d'	ض	d'emeen	ضمان				seezz'	شاذّ
			remed'een	رمضان	13	d	د	deer	دار
			mered'	مرض				medreszeh	مدرسه
5 5	d"	ضّ	fed"ele	فضّل				weled	ولد
			yeh'od"	يحضّ	14	t	ت	toffeeh'	تفّاح
6	z"	ظ	z'eelim	ظالم				kiteeb	كتاب
			mezher	مظهر				nebeet	نبات
			h'eefi'z"	حافظ	15	g	غ	gezeel	غزال
6 6	zz"	ظّ	nezz"eme	نظّم				s'egiir	صغير
			h'ezz"	حظّ				mebleg	مبلغ
7	i	ـِ	e'inszeen	إنسان	16	k	ك	keetib	كاتب
			jibeel	جبال				mekteb	مكتب
			yermii	يرمي				meleek	ملاك
8	e'	.	e'eh'med	أحمد	17	h'	ح	h'emeem	حمام
			moe'min	مؤمن				moh'emmed	محمد
			szemeee'	سماء				mifteeh'	مفتاح
9	o:	ع	o:ereb	عرب	17 17	h"	خ	t'eh"een	طحّان
			moo:ellim	معلّم				s'eh"	صحّ
			yeszmeo:	يسمع	18	c	ث	cewb	ثوب
9 9	o"	عّ	temeo"en	تمعّن				keciir	كثير
			yesio"	يشعّ				celeec	ثلاث

phone-code of that consonant in the phone-code sequences of the input text.

Table 1 illustrates all Arabic symbols and their corresponding phone-codes, phoneme symbols (keyboard symbols). Examples in both symbols are also included in the table showing the possibility of each to be at the beginning, at the end or inside the word.

Table 1-b
Arabic symbols, their corresponding phoneme symbols (Keyboard symbols) and examples in both

Sound code	Symbols Key board	Symbols Arabic	Examples in Keyboard symbols	Examples in Arabic symbols	Sound code	Symbols Key board	Symbols Arabic	Examples in Keyboard symbols	Examples in Arabic symbols
19	m	م	mewz	موز	28	s'	ص	s'ooreh	صوره
			jemiil	جميل				fes'l	فصل
			e'imeem	إمام				miques	مقص
20	n	ن	nehr	نهر	28 28	s"	صّ	xees"eh	خاصه
			e'enee	انا				xes"	خص
			niireen	نيران	29	j	ج	jemel	جمل
21	q	ق	qot'n	قطن				nejjeer	نجار
			bortoqeel	برتقال				ddejeej	دجاج
			t'ebeq	طبق	30	s	ش	sejer	شجر
22	y	ي	yedd	يد				xeseb	خشب
			szeyyideh	سيده				mismis	مشمش
			meyy	مي	31	x	خ	xobz	خبز
23	h	هـ	heetif	هاتف				rexiis'	رخيص
			sehr	شهر				mox	مخ
			mowejjih	موجه	32	l	ل	leyl	ليل
24	w	و	werdeh	ورده				weled	ولد
			lewn	لون				fiil	فيل
			neh'w	نحو	33	r	ر	ref	رف
25	f	ف	fool	فول				moderrisz	مدرس
			xefiif	خفيف				kebiir	كبير
			s'ef	صف	34	o	ـُ	t'oyoor	طيور
26	z	ز	zereefeh	زرافه				yerbit'o	يربط
			mezreo:eh	مزرعه				qors	قرص
			feee'iz	فائز	35	i	ـِ	e'idreb	اضرب
27	sz	س	szemek	سمك				t'ifl	طفل
			e'eszwed	اسود				mis'r	مصر
			moderrisz	مدرس	36	e	ـَ	e'eqeeme	اقام
27 27	ssz	سّ	resszeem	رسام				meqeel	مقال
			h'issz	حس				yeqterih	يقترح

Grapheme-to-phone Code Converter

The grapheme-to-phone-code converter is developed in order to convert the input text into its equivalent phone-codes. This is performed through a tabular chart (tab) in the form of separate rules and some programmable rules. The latter is executed with the main synthesis program. The grapheme of the Arabic input text is in its equivalent phoneme symbols which are shown in *Table 1*. For Hungarian the tab is simple because Hungarian is pronounced almost as it is written and each phoneme has one phone. In contrast to Hungarian, this tab for Arabic is more complicated (1100 rules) because the written version of Arabic is not equivalent to the spoken one. *Fig. 1* shows the conversion process of Arabic input text to phone-codes.

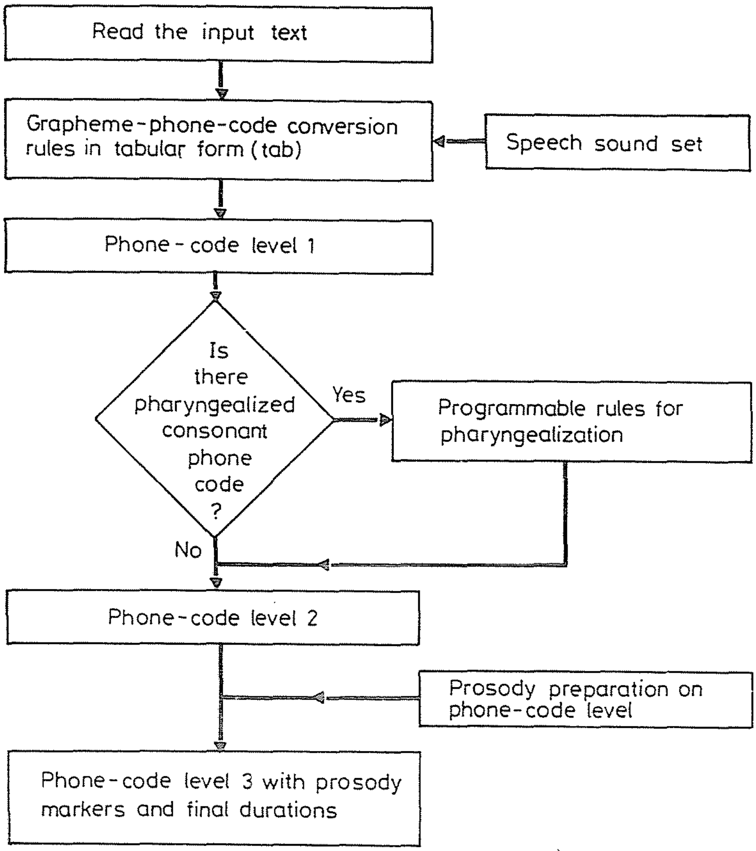

Fig. 1. Flow chart for the conversion of Arabic input text (grapheme) to phone-codes

The syntax of each rule in the tab is written as follows (OLASZY et al., 1990):

Letter(s) and special sign(s)=phone-code(s) and other sign(s)

The grapheme-to phone-code converter table consists of:

1) rules for the conversion of each sound to its own phone-code for all the possible sounds. For example the sounds [t'], [e], [d] and [sz] are converted to their phone-codes by t'=2, 0; e=10; d=13 and sz=27, 0, respectively.

2) rules for solution of the pronunciation problems in the input text. These problems are due to the phonetic differences between the written text and the spoken version of the language. For Arabic the pronunciation problems are very specific and they can be explained in the following:

i) the assimilation process of the sonorant /l/ of the definite article /e'el/ if a definite word begins by sun letter. In Arabic, letters phonemically represented by /t', d", z"n, z', d, t, c, n, z, sz, s', s, l, r/ are traditionally called 'sun letters'. The rest of consonants are not assimilated and they are called 'moon letters'. During the assimilation process, the /l/ of the definite article is not pronounced as /l/ but becomes identical to the following consonant. For example, the word 'the cock' which is represented phonemically by /e'eldiik/ is pronounced as [e'eddiik]. The rule applying to this process is: -e'eld = 1 8 10 14 14 0 Similar rules are written for the other sun letters.

ii) In the pronunciation of two words in which the second one is a definite noun starting the definite article /e'el/ and the first word ending with any short or long vowel, the 'hamza' and its vowel /e'e/ of the definite article are dropped and only the /l/ remains to tie-up the two words. The following two rules illustrate the combination process between the definite article of the second word and the ending of the first word when it is a short vowel and a long vowel, respectively,

e-e'el=10 88 32 0 0 0 0

ee-e'el=10 88 32 0 0 0 0 0

In the above two rules and throughout the paper, the code 88 is not phonecode, but it is a marker for the synthesis program to manipulate the two words separately and to tie-up them only before the synthesizer stage. The same sequence of rules can be used for all the other short and long vowels. If the definite article /e'el/ in the second word is followed by any sun letter, the previous two rules in (ii) are combined with the assimilation rule in (i). In this case, /e'el/ should be dropped and the sun letter is duplicated. The following two rules are devised for a pharyngealized sun letter and for a non-pharyngealized sun letter, respectively:

e-e'elt'=3 88 2 2 0 0 0 0 0

e-e'eld=10 88 13 13 0 0 0 0

In order to illustrate the solution of the above mentioned problems let us consider the following four examples. In the first two examples, we explain the rules mentioned with regard to tie-up two words where the first word ends by a short vowel and the second one starts by the definite article /e'el/ followed by a moon letter or a sun letter. The next two examples describe the cases in which the first word ends by a long vowel and the second word is the same as in the first two examples. *Table 2* illustrates all of these cases.

Table 2
Examples for the utterance of a pair of two words as explained in paragraph ii.

No	English meaning	Phonemical representation	Phonetical representation
1	with the teacher	/meo:e-e'elmoderrisz/	[meo:elmoderrisz]
2	I like the summer	/e'oh'ibbo-e'els'eyf/	[oh'ibbos"ayf]
3	to the school	/e'ilee-e'elmedreszeh/	[ilelmedresze]
4	in the street	/fii-e'elseerio:/	[fisse'rio:]

iii) The third problem in Arabic phonetics is related to the pharyngealized sounds. Pharyngealization is a phenomenon that effects almost all the Arabic speech sounds. The Arabic pharyngealized consonants convert the Arabic vowels [e], [o], [i], [ee], [oo] and [ii] to another group called pharyngealized vowels which are phonetically represented by [a], [u], [i'], [aa], [uu] and [i"], respectively. Many words containing pharyngealized consonant sounds are investigated. It has been found that every pharyngealized consonant has an influence on the neighbouring short or long vowels (up to the third vowel after and down to the second one before). The influence of the pharyngealized consonants on the preceded and followed vowels are solved in the grapheme-to-phone-code converter table by separate rules. The effect of the pharyngealized consonants on the other vowels in the word are performed in programmable rules executed by the main program. Its reason is that these rules would require big size if they were constructed by separate rules in the grapheme-to-phone-code converter table. In addition to these pharyngealized consonant sounds there are two pseudo-pharyngealized consonant sounds; these are [q] and [r]. The uvular [q] is always followed by the pharyngealized vowels [a], [u], [i'], [aa], [uu] and [i']. Also it is always preceded by the vowels [u], [i'] and [ea]. The difference between the pharyngealized consonants and the pseudo-pharyngealized [q] is that the [q] has an influence only on the vowels that precede and follow it. The other pseudo-pharyngealized sound [r] does not always have pharyngealization effect. However, in most words it has pharyngealization effect on the neighbouring short or long vowels (up and down until the second vowel). Still there is an exception sound that has the pharyngealization phenomena only in one word, that is the sonorant [l] in the God's name 'allah'. The [l] is pharyngealized only if 'allah' is pronounced in isolation or preceded by a word ended by the vowel /e/ or /o/. The [l] is non-pharyngealized if 'allah' is preceded by a word ending by the vowel /i/.

iv) Exception words that do not follow the above mentioned rules are put in separate rules. For example, the God's name which is phonemically represented by /e'elleeh/ is written in a separate rule to convert its phonemic representation to the phonetic one [allaah]. This rule is written as follows:

-e'elleeh=1 3 32 32 3 3 23 0 0

Other exception words can be written in a sequence similar to this rule.

Acoustic Building Units (ABUs)

Table 3
Acoustic Building Units (ABUs) for modern standard Arabic

Frame number	Formant frequencies FM1	FM2	FM3	FM4	FM5	Formant baud-widths BW1	BW2	BW3	BW4	BW5	Amp.	Pitch increment	Frame duration
1	433	754	1500	3528	4600	303	126	190	700	800	2	0.0	31.20
2	612	1195	2143	3253	4600	153	68	600	700	800	9	0.0	31.20
3	433	754	1500	3528	4600	303	126	190	700	800	4	0.0	31.20
4	846	2101	2414	3528	4600	153	234	600	700	3000	0	0.0	20.80
5	667	1258	2414	3528	4600	77	234	190	700	3000	0	0.0	10.40
6	217	2450	2719	3528	3900	77	800	600	700	800	9	0.0	31.20
7	398	1996	2414	3253	4600	153	433	190	700	3000	0	0.0	52.00
8	561	1394	2414	3253	4600	153	126	600	700	800	5	0.0	20.80
9	561	1711	2143	2766	3900	153	800	190	700	800	10	0.0	31.20
10	335	1195	2414	3528	4600	153	126	190	700	3000	0	0.0	31.20
11	217	2450	2719	3528	3900	77	800	600	700	800	8	0.0	31.20
12	612	1394	2414	3253	4600	77	126	600	700	3000	0	0.0	10.40
13	433	754	1500	3528	4600	153	126	60	700	3000	0	0.0	10.40
14	667	1626	2414	3528	4600	153	433	190	700	3000	0	0.0	10.40
15	472	1996	2719	3528	4600	153	433	190	700	3000	5	0.0	10.40
16	217	2450	2719	3528	4600	153	800	600	700	3000	0	0.0	10.40
17	398	1626	2414	3253	4600	153	433	190	700	3000	0	0.0	10.40
18	365	1996	2414	3528	4600	153	800	190	265	3000	8	0.0	31.20
19	561	1626	2143	2766	4600	153	800	190	700	3000	0	0.0	10.40
20	258	1195	2414	3528	4600	39	37	60	700	3000	0	0.0	10.40
...	...	...	...	...	...	...	...	...	...	...	...	...	...
255	168	646	2414	3528	4600	153	126	190	700	3000	0	0.0	10.40

At this stage, the input text is converted into appropriate phone-codes. ABUs inventory is developed to represent the basic acoustic level of the language. A special set of ABUs is patented by a Hungarian team in 1985 (OLASZY, 1989). In this paper a set of ABUs for the Arabic language has been constructed. The maximum number of the ABUs and the construction of each one depend on the synthesizer type. However, the contents of each ABU and the number of the ABUs used for the synthesized speech are language dependent. In the case of the free programmable PCF-8200 synthesizer, the maximum number of ABUs that can be generated is 255. In our case for Arabic speech, all the 255 ABUs are used. Each ABU has 8-50 msec duration and consists of formant frequencies, band widths, amplitude, duration and pitch values. The main features of the ABU are that it has a very short duration and it is designed for general purpose use, i.e. it can be adopted and/or interchanged for givin a suitable pronunciation for any sound or sound combinations. The contents of each ABU are adjusted so that it could give a suitable pronunciation for the 36 Arabic sounds and their combinations. *Table 3* includes samples from the ABUs inventory. In fact, the contents of this inventory are obtained from the analysis and the investigation of more than 1000 Arabic words. These words are selected carefully so that all the possible combinations of sounds exist.

PCF-8200 Formant Synthesizer

In the MULTIVOX system the PCF-8200 free programmable formant synthesizer has been used for generating good quality speech from digital code with a programmable bit rate. PCF-8200 is primarily intended for applications in microprocessor controlled systems where the speech code is stored separately. In the synthesizer an excitation signal is fed to a series of resonators. Each resonator simulates one of the formants in the original speech. It is controlled by two parameters, one for the resonant frequency and one for the bandwidth. Five formants are needed for male speech and four for ones female speech. The output of the synthesizer is defined by the excitation signal, the amplitude values and the resonator settings. By periodic updating of all parameters every short time duration (speech frame) very high quality of speech can be produced. Each speech frame constructed for the PCF-8200 synthesizer represents one ABU. The control parameters for each frame (ABU) are in five byte block which contains the filter and source information. The control parameters of the PCF-8200 are illustrated in *Table 4*. The operation of the PCF-8200 formant synthesizer can be explained as follows (PHILIPS, 1986):

1) speech characteristics change quite slowly, therefore the control parameters for the speech synthesizer are adequately updated every few tens of milliseconds with interpolation during the interval in order to ensure a smooth change over from one set of parameter values to the next. In the PCF-8200 the standard frame duration can be set to 8.8, 10.4, 12.8, or

Table 4
Control parameters of PCF-8200 formant synthesizer

Control parameter	Abbreviation	Number of bits
— pitch increment/decrement value	PI	5
— amplitude	AM	4
— frame duration	FD	2
— frequency of 1st formant	FM1	5
— frequency of 2nd formant	FM2	5
— frequency of 3rd formant	FM3	3
— frequency of 4th formant	FM4	3
— frequency of 5th formant	FM5	1
— bandwidth of 1st formant	BW1	3
— bandwidth of 2nd formant	BW2	3
— bandwidth of 3rd formant	BW3	2
— bandwidth of 4th formant	BW4	2
— bandwidth of 5th formant	BW5	2
		40 bits = 5bytes

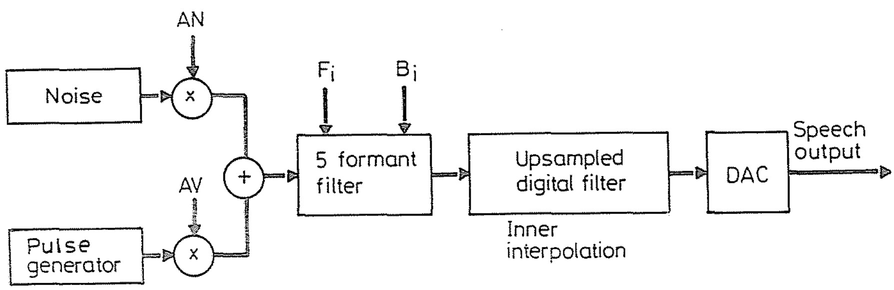

Fig. 2. Block diagram of formant synthesizer

17.6 milliseconds with the speed option, speaking speed, in the command register. The duration of each individual speech frame is programmable to be 1, 2, 3, or 5 times the standard frame duration.

2) the excitation signal is a random noise source for unvoiced sounds and a programmable pulse generator for voiced sounds. Both sources have an amplitude modulator which is updated 8 times in one speech frame by linear interpolation. The pitch is updated every 1/8 of a standard frame.

3) the excitation signal is filtered by five formant filters for male speech and four formant filters for female speech. The formant filter is a cascade of second-order sections. The control parameters, formant frequency and formant bandwidth, are updated eight times per speech frame by linear

interpolation. A block diagram of the formant synthesizer is shown in *Fig. 2.* 4) the filter output is upsampled to 80 kHz and filtered by a digital low-pass filter. Before the signal is digital to analogue converted (DAC), with an 11-bit switched capacitor DAC, the signal is multiplied with a DAC-amplitude factor. The use of a digital filter means that no external audio filtering is required for low-medium applications and minimal filtering is required for those applications where very high quality of speech is needed.

Conclusion

The MULTIVOX, real-time text-to-speech converter system, is adopted for Modern Standard Arabic. A set of 35 sounds for Arabic speech with their all possible phone combinations are defined. A grapheme-to-phone-code converter is developed to convert the input text into its equivalent phone-codes (spoken version). The converter contains solutions for most of the pronunciation problems that may occur due to the differences between the written Arabic text and the spoken version. This has been carried out in the form of separate rules. Solutions for these problems are given in four sets of rules: (i) the pronunciation rules for the sun letters with the definite article /e'el/, (ii) rules for the pronunciation of a pair of two words in which the first word ends with a vowel (short or long) and the second word begins with the definite article /e'el/, (iii) rules relating the pharyngealized vowels to the pharyngealized consonants, and (iv) rules for a set of exception words. The PCF-8200 formant synthesizer has been used and 255 ABUs have been adopted in order to generate good quality of Arabic speech.

Acknowledgements

The authors are grateful to Dr. G. Németh of the Department of Telecommunications and Telematics, Technical University of Budapest, Hungary, for his helpful comments and discussions about the MULTIVOX and to Mr. R. Mrena of the Phonetics Laboratory of the Hungarian Academy of Sciences for his help in the programming of the pharyngealization rules and the preparation of the prosody module routine.

References

EL-IMAM, Y. A. (1987): A Personal Computer Based Speech Analysis and Synthesis Systems,*IEEE-Micro*, pp. 4-21.

EL-IMAM, Y. A. (1989): An Unrestricted Vocabulary Arabic Speech Synthesis System. *IEEE Trans. on Acoustics, Speech and Signal Processing*, Vol. 37, No. 12, pp. 1829-1845.

EL-IMAM, Y. A. (1990): Speech Synthesis Using Partial Syllables. *Computer Speech and Language*, Vol. 4, pp. 203-229.

MOURADI, A.– RAJOUANI, A. – NAJIM, M. (1985): Text-to-Speech Conversion in Arabic Language. *Proc. of Computer Processing of the Arabic Language*, Kuwait, pp. 1-8.

OLASZY, G.– GORDOS, G. (1987): Automatic Text-to-Speech System Applied in a Reading Machine. *Proc. of the European Conf. on Speech Technology*, Edinburgh, Vol. 1, pp. 25-28.

OLASZY, G. (1989): MULTIVOX - A Flexible Text-to-Speech System for Hungarian, Finnish, German, Esperanto, Italian and other Languages for IBM-PC. *Proc. of the European Conf. on Speech Communication and Technology*, Paris, Vol. 2, pp. 525-529.

OLASZY, G. – GORDOS, G. – NÉMETH, G. (1990): Phonetic Aspects of the MULTIVOX Text-to-Speech System. *Proc. of the ESCA Workshop on Speech Synthesis*, Paris, pp. 277-280.

Philips (1986): PCF-8200 Voice Synthesizer: principles and interfacing, Technical Development Data, Netherlands, pp. 407-419.

RAJOUANI, A.– NAJIM, M.– CHIADMI, D. – ZYOUTE, M. (1987): Synthesis-by-rules of Arabic Language", Proc. of the European *Conf. on Speech Technology*, Edinburgh, Vol. 1, pp. 29-32.

SABAH, M. A. – GORDOS, G. – OLASZY G. (1991): A Novel Speech Synthesis Method for Modern Standard Arabic. *Al-Azhar Engineering Second International Conference*, Cairo, Egypt, pp. 141-152.

SELIM, H. – ANBAR, T. (1986): A Phonetic Transcription System of Arabic Text. *IBM Cairo Scientific Center, Technical Report*, pp. 1-27.

Address:

SABAH Mohamed Ahmed,
University of Assiut, Assiut, Egypt.
Géza GORDOS and
Gábor OLASZY
Department of Telecommunications and Telematics
Technical University, H-1521 Budapest, Hungary