PERIODICA POLYTECHNICA SER. EL. ENG. VOL. 36, NO. 2, PP. 93-105 (1992)

DATA-BASE RULE-SYSTEM FOR THE MULTIVOX TEXT-TO-SPEECH CONVERTER APPLICATION FOR ARABIC LANGUAGE

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

The MULTIVOX-Multilingual text-to-speech converter system is adapted to Modern Standard Arabic. In this system, Arabic speech is generated from the concatenation of a set of acoustic building units (ABUs). A 3-dimensional data-base rule-system for the synthesis of unlimited vocabulary Arabic text is organized to concatenate the appropriate ABUs for all possible phone-code pairs that may exist in the input text. The main functions of the MULTIVOX are explained. Illustrative examples are given to show the conversion of Arabic graphemes into phone-codes and the use of the data-base rule-system in the concatenation of the ABUs. Hearing tests have been carried out to test the quality of the synthesized speech.

Keywords: speech synthesis as a rule, data-base rule-systems, text-to-speech systems, Arabic phonetics.

Introduction

Speech synthesis has developed rapidly in the recent years. Speech synthesis as a rule is the most demanding synthesis process with starts with written text as input and produces acceptable speech as output. It is more viable for producing speech from unrestricted text than synthesis from the direct concatenation of pre-recorded units in time domain. This is because the number of the stored units is considerably less in synthesis as a rule than in synthesis by concatenation (YANNAKOUDAKIS, 1987). These rules can be operated on different types of sets of parameters depending on the synthesis technique used. The synthesis from the format parameters (formant synthesis technique) is the most interesting one from the intelligibility of the synthesized speech point of view. For these reasons, synthesis generally based on formant parameters is used in the MULTIVOX.

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

[2] Phonetics Laboratory of the Hungarian Academy of Sciences, Budapest, 1014 P.O. Box 19. Hungary

It is evident that the implementation of the various techniques for speech recognition or speech synthesis is language dependent. Applications to Arabic language are in general not widely known. Among the reasons for this deficiency is the lack of sufficient, reliable, and systematic information about the statistical properties of the structure of both the written and spoken Arabic. The goal of this paper is to provide some basic acoustic information about the Arabic language to narrow the above mentioned gap and to facilitate the development of signal processing technique applications to Arabic, particularly text processing, speech synthesis from written text, speech recognition and others.

This article describes the data-base rule-system for Modern Standard Arabic. These rules include all the possible phone-code pairs for Arabic sounds. Each pair is constructed from a group of Acoustic Building Units (ABUs) according to their serial number in the ABUs inventory (SABAH et al., 1992a). The handling of this rule-system with the ABUs inventory makes the synthesis method more flexible than the diphone segmentation method. This is because, by this method it is easy to change the concatenated ABUs (max. 6) for any phone-code pair in the input text to adjust the required utterance for sound or sound combinations. Also by the possibility of adding the appropriate ABUs, it is easy to adjust the transition between sounds and to deal with the phonetic co-articulation.

MULTIVOX Text-to-Speech System

The MULTIVOX is a multilingual text-to-speech system. It is a patent product of the Technical University of Budapest and the Phonetics Laboratory of the Hungarian Academy of Sciences. The input to the system can be written text in data file or typed text through the keyboard of PC IBM computer of compatible. The output is often good quality speech with intonation and rhythm. It is available now commercially for German, Italian, Hungarian, Dutch, Finnish and Esperanto languages (OLASZY and GORDOS, 1987; OLASZY, 1989; OLASZY et al., 1990). Complete speech synthesis system for Modern Standard Arabic based on the MULTIVOX has been given in (SABAH et al., 1992a). *Fig. 1* illustrates the operation of the MULTIVOX that can be summarized in the following steps.

1) The input text is converted, through grapheme-to-phone-code converter, into sequences of phone-codes. Using this converter some of the pronunciation problems due to the differences between the written text and the spoken version of Arabic speech are solved. The converter and the solutions to the pronunciation problems are explained in (SABAH et al., 1992b).

2) Each phone-code pair is converted to concatenated ABUs through a matrix rule system. During this process a prosody module has been adopted to take into consideration the intonation and the rhythm of the given input text.
3) The output of the matrix rule system that represent the acoustic level of the input text is coded in 5 bytes control codes. These codes are transferred to the PCF 8200 speech synthesizer.
4) The output of the synthesizer is rhythmic good quality speech.

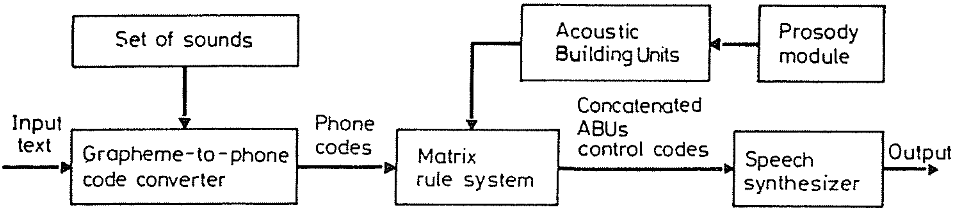

Fig. 1. Block Diagram of the MULTIVOX

Rule System

The generation of synthesized speech requires, in general, the construction of a phonetic dictionary to include the phonetic segments (units) that represent the acoustic part of the language and a set of rules to concatenate these units is necessary. In fact, the direct concatenation of the speech units usually produces poor quality speech (unintelligible). To improve the quality complex rules are required to take into consideration the co-articulations and transitions between sounds. The construction of rules to link and match the phonemic representation of the input text with appropriate speech segments requires and extensive knowledge and deep understanding of the speech production of the language. In the Arabic version of the MULTIVOX, very short time duration speech segments (ABUs) have been introduced. Each ABU can be adapterd or interchanged to solve the problems of co-articulations and transition between sounds (SABAH et al, 1992b).

Arabic alphabet consists of 29 letters. Among them 28 are consonants and one long vowel. In some sequences, two from the consonants serve as long vowels. Also, there are three short vowels in the Arabic language. 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

Table 1-a
Consonants, IPA Symbols, Phoneme and Phone Codes

Arabic Letter	letter name	IPA symbol	Phoneme* symbol	Phone code	Examples in phoneme symbols
ط	(ta)	ṭ	[t']	2	met'eer
ض	(dad)	ḍ	[d']	5	remed'een
ظ	(thah)	∂	[z"]	6	z"eelim
ء	(hamza)	?	[e']	8	moe'min
ع	('ain)	ʕ	[o:]	9	o:ereb
ب	(beh)	b	[b]	11	beled
ذ	(thel)	∂	[z']	12	z'eheb
د	(dal)	d	[d]	13	deer
ت	(teh)	t	[t]	14	kiteeb
غ	(ghain)	ɡ	[g]	15	gezeel
ك	(kaf)	k	[k]	16	keetib
ح	(ha)	h	[h']	17	h'emeem
ث	(theh)	Θ	[c]	18	cemen
م	(mîm)	m	[m]	19	jemiil
ن	(nûn)	n	[n]	20	nehr
ق	(qaf)	q	[q]	21	qelem
ي	(yeh)	y	[y]	22	yeneem
ه	(heh)	h	[h]	23	heetif
و	(waw)	w	[w]	24	weled
ف	(feh)	f	[f]	25	xefiif
ز	(zîn)	z	[z]	26	zeyt
س	(sîn)	s	[sz]	27	szemek
ص	(sad)	S	[s']	28	fes'l
ج	(jîm)	J	[j]	29	jemel
ش	(shîn)	ʃ	[s]	30	xeseb
خ	(kha)	x	[x]	31	xobz
ل	(lam)	l	[l]	32	o:elem
ر	(ra)	r	[r]	33	reszeme

carrying them. These short vowels are in contrast phonemically with the three long vowels. The consonants and the short vowels are represented by 31 phonemes. For the purpose of the present system, the 31 phonemes are introduced and they are converted by the grapheme-to-phone-code converter into 36 phone-codes. From these phone-codes, the phone-code 1 is devoted for the space that often occurs at the beginning of each word, and 28 phone-codes are devoted to the consonants where each phoneme has one phone-code. The other seven phone-codes are devoted to the short vowels. In the system each long vowel is represented by doubling the phone-code of the corresponding short vowel. In addition, long vowels are considered to be close to the double duration of their corresponding short ones. *Table 1* illustrates the Arabic consonants and vowels, their IPA symbols and their phoneme symbols together with one example for each sound. Since the pronunciation of any sound depends on the preceding and the following sounds, the possible sound combinations that exist in Arabic speech are defined. Although some of these combinations do not exist (e.g. the sound of [s'] followed by [sz]), rules are designed for such combinations since they may exist in foreign words.

Table 1-b

Vowels, IPA Symbols, Phonemes, Phones and Phone Codes

Arabic vowel	Vowel name	Vowel type	IPA symbol	Phoneme symbol	Phone symbol	Phone code		Examples in phoneme symbols
"ـَ"	elfetheh	short	a	e	[3]	3		qelem
					[e]	10		e'emel
					[ea]	**36**		meqeel
"ـُ"	eldammah	short	u	o	[o]	4		doleeb
					[u]	**34**		qot'n
"ـِ"	elkasrah	short	i	i	[i]	7		jibeel
					[i']	**35**		t'ifl
ا	mad beelalif	long	a:	ee	[aa]	**3**	**3**	t'eee'ir
					[ee]	10	10	meee'il
					[eaea]	36	36	not used
ـُو	mad beelwaw	long	u:	oo	[oo]	4	4	fool
					[uu]	**34**	**34**	s'ooreh
ـِي	mad beelyaa	long	i:	ii	[ii]	7	7	jemiil
					[i"]	**35**	**35**	s'egiir

In the system a grapheme-to-phone-code converter is designed to convert the input text to a series of phone-codes depending on extensive phonetic rules for Arabic speech. Then, a data-base rule-system is developed for the segmentation of these phone-codes to phone-code pairs and to represent each phone-code pair by a set of concatenated ABUs. To perform this process a three dimensional matrix (X.X.6) is organized to represent the rule system for the MULTIVOX. The (X.X) matrix defines all the possible phone-code pairs in the language. The third dimension (6) represents the maximum number of the concatenated ABUs required for each phone-code pair. For Arabic X stands 36 including the phone-code 1 that represents the pause at the beginning of each word. The output of this matrix is a series of ABUs, each has its own serial number in the inventory. These ABUs

Table 2-1

Rules for the combinations of phones [-] and [t'] that have phone-codes 1 and 2, respectively with the other Arabic phone

combinations of [-]

Rule number	Rule					
(1, 1)	40	0	0	0	0	0
(1, 2)	40	46	46	0	0	0
(1, 3)	12	2	22	0	0	0
(1, 4)	13	3	23	0	0	0
(1, 5)	208	49	49	0	0	0
(1, 6)	158	25	25	78	0	0
(1, 7)	16	6	26	0	0	0
(1, 8)	17	7	0	0	0	0
(1, 9)	14	231	224	0	0	0
(1, 10)	19	9	29	0	0	0
(1, 11)	32	32	32	39	0	0
(1, 12)	158	170	177	0	0	0
(1, 13)	208	49	80	56	0	0
(1, 14)	46	46	71	0	0	0
(1, 15)	103	102	42	42	0	0
(1, 16)	40	10	46	0	0	0
(1, 17)	183	194	194	0	0	0
(1, 18)	41	41	67	71	74	0
(1, 19)	149	142	142	0	0	0
(1, 20)	149	152	150	0	0	0
(1, 21)	93	213	219	0	0	0
(1, 22)	16	189	26	0	0	0
(1, 23)	225	226	226	0	0	0
(1, 24)	149	143	2325	146	0	0
(1, 25)	131	132	0	0	0	0
(1, 26)	158	173	173	92	0	0
(1, 27)	191	165	164	0	0	0
(1, 28)	191	140	179	0	0	0
(1, 29)	99	101	95	0	0	0
(1, 30)	182	180	0	0	0	0
(1, 31)	232	234	230	0	0	0
(1, 32)	215	196	195	0	0	0
(1, 33)	220	212	216	212	0	0
(1, 34)	13	241	241	0	0	0
(1, 35)	158	135	0	0	0	0
(1, 36)	40	125	125	0	0	0

combinations of [t']

Rule number	Rule					
(2, 1)	72	85	57	162	30	0
(2, 2)	10	10	46	0	0	0
(2, 3)	71	83	57	2	22	0
(2, 4)	73	71	157	34	0	0
(2, 5)	73	71	157	162	49	50
(2, 6)	73	71	157	25	25	78
(2, 7)	73	71	157	189	6	0
(2, 8)	157	71	72	162	17	7
(3, 9)	73	72	157	162	8	224
(2, 10)	73	71	157	29	9	0
(2, 11)	72	71	157	162	32	32
(2, 12)	157	73	71	170	177	0
(2, 13)	157	71	73	157	49	50
(2, 14)	73	72	162	46	46	0
(2, 15)	73	71	157	103	42	44
(2, 16)	73	72	162	46	46	0
(2, 17)	73	74	157	162	194	194
(2, 18)	71	73	157	20	20	74
(2, 19)	73	72	220	149	142	142
(2, 20)	73	72	220	150	150	0
(2, 21)	73	72	162	219	0	0
(2, 22)	73	72	251	136	185	0
(2, 23)	73	71	157	225	226	0
(2, 24)	73	72	220	149	145	121
(2, 25)	73	72	162	132	132	0
(2, 26)	73	72	162	173	92	0
(2, 27)	73	72	41	165	164	0
(2, 28)	73	72	41	140	179	0
(2, 29)	73	72	157	162	101	95
(2, 30)	157	73	182	180	0	0
(2, 31)	73	71	232	234	230	0
(2, 32)	73	72	157	215	195	195
(2, 33)	73	72	157	212	211	212
(2, 34)	73	71	157	241	241	0
(2, 35)	73	72	220	171	135	0
(2, 36)	73	71	157	125	125	0

Table 2-2

Rules for the Combinations of Phones [a] and [o] that have phone-codes 3 and 4 respectively with the other Arabic phone

combinations of [a]

Rule number	Rule					
(3, 1)	22	22	12	30	0	0
(3, 2)	22	46	46	10	0	0
(3, 3)	2	2	22	22	0	0
(3, 4)	22	23	3	0	0	0
(3, 5)	22	49	49	80	80	0
(3, 6)	22	25	25	78	0	0
(3, 7)	22	26	26	6	0	0
(3, 8)	22	17	7	27	0	0
(3, 9)	2	8	231	231	0	0
(3, 10)	22	29	9	0	0	0
(3, 11)	22	39	39	0	0	0
(3, 12)	2	158	177	0	0	0
(3, 13)	22	49	80	0	0	0
(3, 14)	22	46	46	255	0	0
(3, 15)	22	103	102	42	44	0
(3, 16)	22	46	46	0	0	0
(3, 17)	156	200	194	194	0	0
(3, 18)	22	46	46	71	71	74
(3, 19)	22	149	146	142	0	0
(3, 20)	22	155	150	150	0	0
(3, 21)	22	213	219	219	0	0
(3, 22)	22	138	136	136	0	0
(3, 23)	22	226	226	0	0	0
(3, 24)	22	167	128	130	0	0
(3, 25)	22	132	132	132	0	0
(3, 26)	22	173	173	0	0	0
(3, 27)	22	166	222	164	164	0
(3, 28)	22	140	165	179	179	0
(3, 29)	2	99	111	95	0	0
(3, 30)	22	166	180	180	0	0
(3, 31)	22	234	230	0	0	0
(3, 32)	22	215	195	195	0	0
(3, 33)	22	220	212	211	212	0
(3, 34)	22	241	242	0	0	0
(3, 35)	22	135	135	0	0	0
(3, 36)	22	125	125	0	0	0

combinations of [o]

Rule number	Rule					
(4, 1)	23	13	30	0	0	0
(4, 2)	34	46	46	46	0	0
(4, 3)	23	2	22	0	0	0
(4, 4)	1	1	34	34	0	0
(4, 5)	34	49	49	68	0	0
(4, 6)	23	25	25	78	0	0
(4, 7)	23	26	6	0	0	0
(4, 8)	23	13	17	7	0	0
(4, 9)	23	8	231	0	0	0
(4, 10)	23	29	9	0	0	0
(4, 11)	23	32	32	36	0	0
(4, 12)	23	158	170	170	177	0
(4, 13)	23	49	80	0	0	0
(4, 14)	23	46	46	255	0	0
(4, 15)	3	23	104	42	44	0
(4, 16)	23	46	46	0	0	0
(4, 17)	23	58	89	194	194	0
(4, 18)	23	46	71	71	74	0
(4, 19)	1	146	142	142	0	0
(4, 20)	34	150	150	0	0	0
(4, 21)	23	213	219	219	0	0
(4, 22)	34	137	136	0	0	0
(4, 23)	34	226	226	0	0	0
(4, 24)	23	149	127	0	0	0
(4, 25)	23	132	132	132	0	0
(4, 26)	3	173	0	0	0	0
(4, 27)	167	191	164	164	0	0
(4, 28)	23	191	140	179	0	0
(4, 29)	23	104	101	102	95	0
(4, 30)	23	13	167	180	0	0
(4, 31)	23	230	237	0	0	0
(4, 32)	23	215	195	195	0	0
(4, 33)	23	34	212	211	212	0
(4, 34)	23	3	241	0	0	0
(4, 35)	23	3	135	135	0	0
(4, 36)	23	125	125	0	0	0

represent the control codes of the synthesizer. *Table 2* shows some of the 36.36 rules with the appropriate ABUs for each rule for all the phone-code pairs of Arabic sounds. A complete list of these rules can be obtained from the authors.

Illustrative Examples

Three examples are given to illustrate the conversion process and the appropriate ABUs concatenated for each phone-code pair. *The grapheme of the Arabic input text is represented by its equivalent phoneme symbols given in Table 1.* Also, the pharyngealized vowels ([a], [u], [i'], [aa], [uu], [i"]) and the vowel [ea] are not supported in the input text, however, they occur only in the phone-code sequences according to rules developed in the grapheme-to-phone-code converter table and programmable rules executed with the main synthesis program.

Example (1): This example shows the conversion process of the word 'fish' that has the same phonemic and phonetic representations. It is represented phonemically by (szemek). Consequently the conversion is direct, i.e., each phoneme is converted to its own phone-code. *Table 3* illustrates the conversion process of this word.

Example (2): In this example the conversion process of the word 'spring' will be presented. The phonemic representation of this word is (rebiio:), but this representation is not compatible with the phonetic one. We show that this word contains the trill [r], and this sound is often followed and preceded by the vowel sound [a], not by [e] except in some exception words. Rules were designed in the grapheme-to-phone-code converter to perform this process, and to yield the suitable phone-codes (SABAH et al., 1992b). Therefore the appropriate vowel sound for this word is [a], with phone-code 3, instead of the vowel sound [e]. *Table 4* shows the conversion process for the word 'spring'.

Example (3): In this example the conversion process of the word 'children' will be presented. This word is phonemically represented by (e'et'feel), but this representation is not compatible with the phonetic one because it contains the pharyngealized consonant sound [t']. This sound is always followed and preceded by one of the pharyngealized vowels. Rules are designed in the grapheme-to-phone-code converter table to take into consideration the effect of the pharyngealized consonant on the following and preceding phoneme vowels. Consequently, the short vowel [e] before the consonant [t'] in the word (e'et'feel) is replaced by the vowel sound [a]. Because the glottal sound [e'] with the vowel sound [a] is not phonetically compatible at the beginning of the word, the combination [e'a] is replaced

Table 3
Conversion Process from Grapheme into Concatenated ABUs when the Word 'fish' with Phonemic Symbols (szemek) is synthesized

Phoneme (grapheme)	Phone code	Matrix element	Equivalent ABUs (contents of the matrix element)	Meaning of ABUs
–	1	(1, 27)	40 191 165 164 0 0	191 165 & 164 are the parts of [sz]
sz	27	(27, 10)	164 166 9 9 0 0	9 & 9 are the parts the of [e]
e	10	(10, 19)	29 149 142 142 0 0	149 142 & 142 are the parts of [m]
m	19	(19, 10)	142 29 29 0 0 0	29 & 29 are the parts of [e]
e	10	(10, 16)	29 41 41 41 0 0	41 41 & 41 are the silent parts of [k]
k	16	(16, 1)	20 119 120 30 0 0	120 and 119 are the voiceless parts of [k]
–	1			

only by [a] with phone-code 3. While the effect of a pharyngealized sound on the other vowels (the second and/or the third) in a word is developed on the basis of programmable rules to execute with the main synthesis program. According to one of to the programmable rules, the long vowel [ee] in (e'et'feel) is converted to the pharyngealized counterpart [aa] with phone-code 3,3. As a result, this word is phonetically represented by [at'faal]. *Table 5* illustrates the conversion process of this word.
In *Tables 3*, *4* and *5*, the speech frames 40 and 30 are the acoustic representation of the beginning and the ending of the word, respectively.

Table 4
Conversion Process from Grapheme into Concatenated ABUs When the Word 'spring' with Phonemic Symbols (rebiio:) is synthesized

Phoneme (grapheme)	Phone code	Matrix element	Equivalent ABUs (contents of the matrix element)	Meaning of ABUs
–	1			
		(1, 33)	40 212 216 212 0 0	212, 216 and 212 are the parts of [r]
r	33			
		(33, 3)	211 212 211 2 22 0	2 and 22 are the parts of [a]
e	3			
		(3, 11)	22 39 39 0 0 0	39 & 39 are the silent parts of [b]
b	11			
		(11, 7)	36 33 35 35 0 0	36 & 33 are the voiced parts of [b]
i	7			
		(7, 7)	118 118 118 118 0 0	35 & 118 are the parts of [ii]
i	7			
		(7, 9)	26 4 8 231 0 0	4, 8 & 231 are the parts of [o:]
o:	9			
		(9, 1)	224 236 8 14 30 0	224 236 8 14 are the end parts of [o:]
–	1			

Evaluation of the Synthesized Speech

To evaluate the speech quality of the Arabic version of the MULTIVOX, synthetic speech representing different words and sentences is created and hearing test is performed. The goal of this test is to evaluate the intelligibility and naturalness of the synthetic output speech. The subjects on which the test is carried out are two 6-person groups (group 1 and group 2) who are native speakers of Arabic, both with no history with the content of the input text. Group 1 consists of persons who have previous experience with

Table 5
Conversion Process from Grapheme into Concatenated ABUs When the Word 'children' with Phonemic Symbols (e'et'feel) is synthesized

Phoneme (grapheme)	Phone code	matrix element	Equivalent ABUs (contents of the matrix element)						Meaning of ABUs
–	1								
		(1, 3)	40	12	2	22	0	0	12, 2 and 2 are the parts of [a]
e'	}→ 3								
e									
		(3, 2)	22	46	46	10	0	0	46, 46 and 10 are the silent parts of [t']
t'	2								
		(2, 25)	73	72	162	132	132	0	73 and 72 are the voiceless parts of [t']
f	25								
		(25, 3)	132	22	2	0	0	0	132 is for [f]
e	3								
		(3, 3)	2	2	22	22	0	0	2, 22 and 22 are for [aa]
e	3								
		(3, 32)	22	215	195	195	0	0	215, 195 & 195 are the parts of [l]
l	32								
		(32, 1)	195	254	215	30	0	0	254, 215 and 30 are the end parts of [l]
–	1								

the synthesized speech resulting from the system during the developmental phase. Group 2 consists of persons with no previous experience with the synthesized speech.

The test materials consist of about 120 isolated words, 20 general sentences contain all the different prefixes, prepositions, conjunctions and articles, and 23 question sentences contain most Arabic question words. The reading module of the system has been used to read automatically the text materials from data files. The material under test contains the representative from all the classes of Arabic speech sounds, phonetic problems that may be contained in any text, stressed and unstressed words. The subjects listen to the synthetic output speech directly from the headphone of the system. In most cases a success of recognition is achieved from the first trial. In cases where there is a failure in the recognition of any word or sentence from the first trial, repetitive trials (limited to three) are made. The text of the synthesized speech is then provided to the subjects and they are asked to indicate all synthetic units or phonemes that are not intelligible to them. *Tables 6* and *7* show the results of the hearing test for sentences and words, respectively. The results of the hearing test show that the intelligibility score for sentences is higher than that of isolated words. Statistical analysis shows that the intelligibility scores for isolated words and sentences are 76 % and 87 %, respectively.

Table 6

Results of sentences intelligibility test

Group number	Sentence type	Subjects	1	2	3	4	5	6	Overall mean
1	question	grand mean	95.1	89.1	90.0	90.2	91.0	92.0	91.2 %
	declarative	grand mean	90.7	86.1	84.8	85.0	90.2	87.0	87.3 %
2	question	grand mean	87.5	89.4	88.0	89.8	92.4	88.1	89.3 %
	declarative	grand mean	82.6	80.7	79.1	83.1	81.3	86.4	82.2 %

Table 7

Results of word intelligibility test

Group number	Position of sound	Subjects	1	2	3	4	5	6	Overall mean
1	word initial	grand mean	82.5	80.0	79.3	84.8	78.9	81.0	81.1 %
	word final	grand mean	79.8	74.0	76.6	78.8	75.0	77.0	76.9 %
	word middle	grand mean	82.8	78.8	79.5	80.6	81.3	80.0	80.5 %
	doubled sound	grand mean	83.5	75.5	83.5	72.7	81.2	85.0	80.2 %
2	word initial	grand mean	74.4	77.4	76.0	75.2	71.5	79.0	75.6 %
	word final	grand mean	67.3	66.6	65.6	71.2	69.5	76.1	69.4 %
	word middle	grand mean	70.6	74.4	71.2	76.8	72.8	78.9	74.1 %
	doubled sound	grand mean	74.1	73.5	74.1	80.9	70.6	81.6	75.8 %

Conclusion

A data-base rule-system for the MULTIVOX, real-time text-to-speech converter, is adopted for Modern Standard Arabic. The system deals with most of the phonetic problems in Arabic speech. The philosophy of the used synthesis method depends on the fact that the speech can be formed by the concatenation of a set of ABUs. A 3-dimensional matrix data-base rule-system is organized to concatenate the appropriate ABUs for all the possible phone-code pairs that may exist in the input text. Illustrative examples are included to explain the conversion process and the use of these rules in the synthesis of three Arabic words. Finally, hearing test is carried out to evaluate the quality of the synthesized speech. It shows that the intelligibility of the resulting speech is highly accepted.

References

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.

SABAH, M. A. – GORDOS, G. – OLASZY, G. (1992a): Unlimited Vocabulary Speech Synthesis System for Modern Standard Arabic. Submitted to *Journal of Computer Speech and Language.*

SABAH, M. A. – GORDOS, G. – OLASZY, G. (1992b): Acoustic Building Units for Formant Synthesis Text-to-Speech Converter System for Modern Standard Arabic. Submitted to the *Periodica Polytechnica,* Technical University of Budapest.

YANNAKOUDAKIS, E. J. – HUTTON, P. J. (1987): Speech Synthesis and Recognition Systems. John Wiley and Sons, New York.

Address:

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