A FULL HUNGARIAN TEXT-TO-SPEECH MICROCOMPUTER FOR THE BLIND

G. Kiss(1), A. Arató(2), J. Lukács(3), J. Sulyán(2), T. Vaspöri(2)

1. Hungarian Academy of Sciences Institute of Linguistics
2. Hungarian Academy of Sciences Central Reas.Inst.for Physics
3. Eötvös Loránd Univ. Faculty of Natural Sciences

ABSTRACT

The authors introduce Braille-Lab, a Hungarian-speaking microcomputer developed for the blind. This Z80 microprocessor-based personal computer is fitted with a Philips MEA 8000 formant synthesizer, providing for Hungarian text-to-speech conversion. The original version of the machine contains a speaking BASIC interpreter. The new version, Braille-Lab+, is also furnished with a speaking word processor and a speaking database management system running under a speaking CP/M compatible operating system. Braille-Lab has been approved and adopted by the Hungarian National Federation of the Blind, 95 sets have been installed so far.

INTRODUCTION

In the past few decades, intensive research into speech synthesis has been going on in a number of countries including Hungary. This research work has three main types of motivation.

1. Fifth-generation computers are to create a new, humanized type of man-machine-man relationship. Hence one of the main objectives of research is viva voce 'conversation' between man and machine. The various links of the man-machine-man communication chain (each constituting a research area in its own right) and the way artificial speech production fits into that chain are represented in Fig. 1.

2. Another impulse for attempts at speech synthesis was the desire to achieve a better understanding of the acoustics of speech. Indeed the principle of analysis by synthesis is more effective than any measuring apparatus, however sophisticated the latter may be: it shows what the essential components of speech really are [7]. That principle can be best implemented by formant synthesis. These considerations led to the establishment, under Kálmán Bolla's leadership,of a complex acoustic

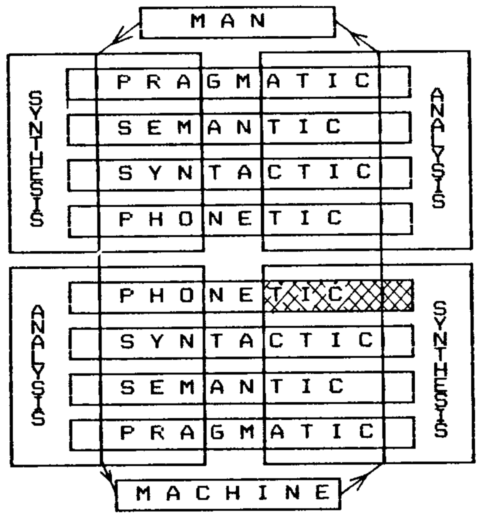


Fig. 1. The Man–Machine–Man communication chain

speech synthesizing system in the Linguistics Institute of the Hungarian Academy of Sciences, in the late 1970s. The hardware configuration includes an OVE III (Swedish-made) formant synthesizer [1] and a PDP 11/34 computer. The effective operation of the system is guaranteed by a specially designed interactive program called FOPRO [11]. The utility of the system for phonetic research is demonstrated by a number of scholarly papers [5, 10]. The program was also used for designing an inventory of speech frames for a Hungarian text-to-speech (TTS) system based on the principle of formant synthesis in the early 1980s [13]. The inventory, in turn, was used in HUNGAROVOX, a Hungarian real-time TTS system for speech synthesis [9, 12]. Later, a developing system was also made for a Philips MEA 8000 formant synthesizer [3].

3. The third type of motivation for research on speech synthesis is a desire to develop various appliances to help handicapped people (afflicted with speech disorders, blindness, etc.). The area was

given a vast impetus by the appearance, in the early 1980s, of speech synthesizer contained in a single IC, e.g. UAA 1003, TMS 5200, SC-01, MEA 8000 [2, 4, 6, 8], since these could be built into various appliances. These considerations led to the development of Braille-Lab (B-L), a speaking computer to be used by blind people, introduced in the present paper. This Hungarian-speaking microcomputer fitted with a text-to-speech conversion system effectively helps the education of blind people in computational technology (thus creating high-qualification employment possibilities for them) . Also, it accelerates their full integration into society.

THE HARDWARE OF BRAILLE-LAB

B-L is a Hungarian-made, Z80 microprocessor-based personal computer. Its memory is organized on a page basis,and consists of 64 kbyte RAM and 20 kbyte ROM. The card containing the speaking module has been built into the computer with MEA 8000. The TTS software is located on page 2 of ROM. The keyboard of B-L contains every letter of the Hungarian alphabet, arranged in a way almost identical with the keyboard of standard Hungarian typewriters. The built-in small loudspeaker makes it possible for the speech produced by the system to be heard without an external loudspeaker. The built-in BASIC interpreter leaves 48 kbyte free memory capacity available for the user.

The basic version of B-L has been futher developed. B-L+, the new version, runs under a CP/M compatible operating sysem. Along with a 64 kbyte operative memory, it is also furnished with a 192 kbyte RAM disk and a 1 Mbyte floppy disk drive. The new version further contains a speaking word processor and a speaking database management system. With these two programs,its possibilities of application by the blind have been multiplied.

THE TEXT-TO-SPEECH SOFTWARE SYSTEM OF BRAILLE-LAB

The basis for Hungarian TTS conversion by B-L is a text in Hungarian orthography, with no special symbols added. The program translates that text into a series of frame code numbers for the MEA 8000 synthesizer. The frame code numbers designate the elements of a 218-member frame inventory, devised earlier. The TTS conversion is implemented in the following four steps:

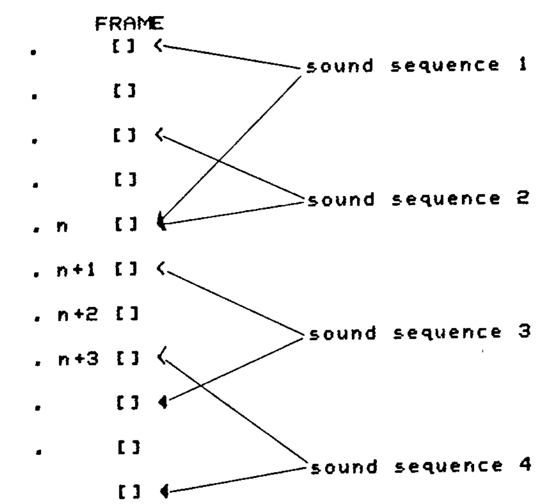
1. First of all, the text to be converted to speech is transformed by the program into a series of (code numbers of) speech sounds. Hungarian orthography is a fairly accurate indicator of the series of sounds to be uttered. However, not only single letters but also combinations of two, and even three, letters may stand for single sounds. In the letter-to-sound transformation, the program basically relies on Fig. 2.:

1	2	3	4	5	1	2	3	4	5
1.	a	1	ɔ	-	34.	nn	22	n:	+
2.	á	2	a:	-	35.	ny	23	ɲ	-
3.	b	10	b	-	36.	nny	23	ɲ:	+
4.	bb	10	b:	+	37.	o	6	o	-
5.	c	11	t͜s	-	38.	ó	6	o:	+
6.	cc	11	t͜s:	+	39.	ö	7	ø	-
7.	cs	12	t͡s	-	40.	ő	7	ø:	+
8.	ccs	12	t͡s:	+	41.	p	24	p	-
9.	d	13	d	-	42.	pp	24	p:	+
10.	dd	13	d:	+	43.	r	25	r	-
11.	e	3	ɛ	-	44.	rr	25	r:	+
12.	é	4	e:	-	45.	s	26	ʃ	-
13.	f	14	f	-	46.	ss	26	ʃ:	+
14.	ff	14	f:	+	47.	sz	27	s	-
15.	g	15	g	-	48.	ssz	27	s:	+
16.	gg	15	g	+	49.	t	28	t	-
17.	gy	16	ɟ	-	50.	tt	28	tt:	+
18.	ggy	16	ɟ:	+	51.	ty	29	c	-
19.	h	17	h	-	52.	tty	29	c:	+
20.	hh	17	h:	+	53.	u	8	u	-
21.	i	5	i	-	54.	ú	8	u:	+
22.	í	5	i:	+	55.	ü	9	y	-
23.	j	18	j	-	56.	ű	9	y:	+
24.	jj	18	j:	+	57.	v	30	v	-
25.	k	19	k	-	58.	vv	30	v:	+
26.	kk	19	k:	+	59.	z	31	z	-
27.	l	20	l	-	60.	zz	31	z:	+
28.	ll	20	l:	+	61.	zs	32	ʒ	-
29.	ly	18	j	-	62.	zzs	32	ʒ:	+
30.	lly	18	j:	+	63.	sp	33	-	-
31.	m	21	m	-					
32.	mm	21	m:	+					
33.	n	22	n	-					

1= number, 2= letter /s/, 3= code number
4= IPA symbol , 5= length of sound

Fig. 2. Table of letter-to-sound correspondences

2. The second step of TTS conversion is the designation of the series of frames that will realize the speech sounds of the text to be uttered. This designation is basically of a diadic nature.The 218 frames utilized are arranged in the inventory in a very special order. Each combination of sounds is realized by adjacent frames. Thus we can dispense with storing what is called a combination matrix and consequently save a significant amount of memory capacity. In order to further optimize the utilization of the frame inventory, various sound sequences can be realized by overlapping series of frames,as illustrated in Fig. 3. Long sounds are also produced at this stage by multiplying some component of the frame of the corresponding short sound (2 to 5 times, as the case may be) in the series of frame code numbers. Each element of the series of frame code numbers will be an integer between 1 and 218. That series then serves as input to the melody generating part of the program.



<--initial element of series of freme
◄-- final element of series of frame

Fig. 3. The structure of the frame inventory and the way frames realizing sound sequences are specified

3. The melody is generated by the program by selecting the appropriate value of the PI parameter of the MEA 8000 synthesizer frame-by-frame. The first step in producing the melody is the segmentation of the text into intonation units. The intonation units are marked off by .(full stop) ,(comma) ?(question mark) !(exclamation mark)or RETURN. Triggered by those punctuation marks, the program will supply the segmental structure produced so far with one of the melody patterns.

4.When the coding of segmental and suprasegmental structure is completed, B-L forwards the resulting series of code numbers to the MEA 8000 speech synthesizer, the speech is simultaneously heard.

THE USE OF BRAILLE-LAB

The computer is able to speak as soon as it is switched on. The following introductory words appear on the screen and are simultaneously heard [in Hungarian]: "B-L computer, version 3.1. 48843 bytes of free memory capacity."

After that, each time a key is pressed, the system utters the corresponding speech sound, in order to make it easier for a blind person to avoid typing errors. Names of non-letter keys, including numerals, are uttered as words. E.g. on pressing % the machine says "százalék" (percent), etc. Using the cursor keys, the user can aurally check the contents of any character position of the screen.

Basically there are two situations in which B-L actually speaks: 1. during entering and editing BASIC programs; 2.at run-time when any information appearing on the sreen is simultaneously said aloud.

1. During program editing, the echoing function mentioned above is in operation; in addition, at the end of each line when RETURN is pressed the computer reads out the whole line as connected text. Numerals at this point are not read character-by-character but as wholes (e.g. twenty-five rather than two, five). The English terms of BASIC are read out according to the Hungarian value of

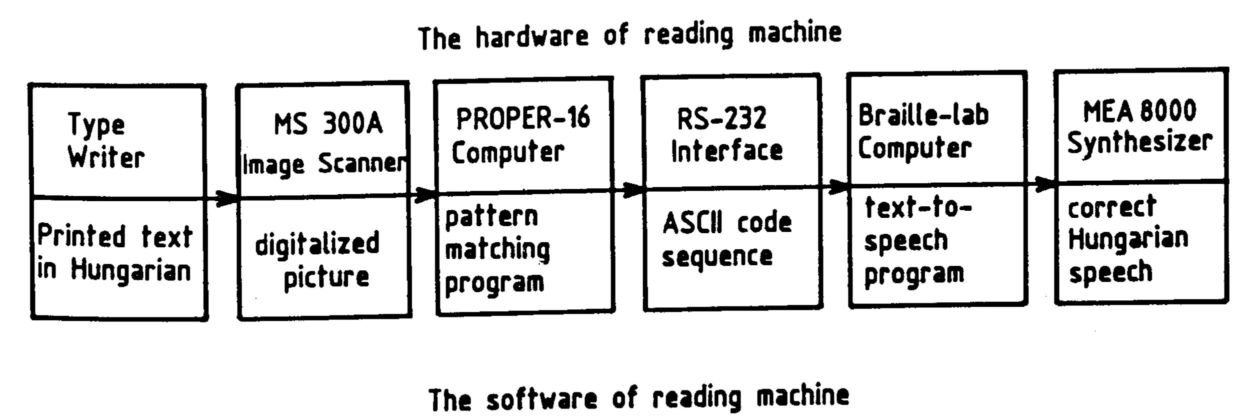


Fig. 4 The system of the reading machine when the Braille-lab is speaking terminal

letters, rather than in proper English pronunciation. At program listing, the list can be heard as it appears on the screen. In short, any information appearing on the screen including e.g. error messages, is also uttered without any special command.

2. The information appearing on the screen during the running of BASIC program will also be heard automatically. For instance, as a result of the running of the following short program, all Hungarian numerals between 1110 and 1125 will be heard first with a question intonation and then with a statement intonation (i.e. "Is the next number 1110? Yes, 1110." etc):

```
10 FOR I=1110 TO 1125
20 PRINT "A következő szám" I "?"
30 PRINT "Igen" I "."
40 NEXT
```

BRAILLE-LAB AS THE SPEAKING PERIPHERY OF A READING MACHINE

At an exhibition called "Hungarians in the World" held in Budapest in August 1986, the authors, assisted by researchers of SZKI (Institute for Computer Research) connected B-L whith an IBM compatible PROPER-16 computer. On the other side, an MS 300A Image Scanner was also connected with PROPER-16. The image recognition program developed by the SZKI people recognized printed Hungarian text. PROPER-16 forwarded the resulting ASCII code, via a standard RS 232 interface, to B-L which uttered the text real time, with a proper Hungarian intonation, intelligibly.

HOW TO MAKE BRAILLE-LAB SING

One of the special features of B-L is that it can also sing. To make the computer sing, the user has to specify the correct rhythm and the correct sequence of pitches. Rhythm can be represented by lengthening the vowels appearing in the words of the song, by entering vowel letters more than once. The length of the syllable containing the vowel will increase in proportion with the number of identical vowel letters entered. The melody has to be given in relative sol-fa letters, according to Zoltán Kodály's method. The pitch defined by sol-fa letter assigned to a syllable will be superimposed by the program on the appropriate syllable which has been rhythmically defined as above. By that procedure, any Hungarian-text song can be produced. This special feature of the system opens up a novel area of application in the on-line representation and correction of Braille music notation [15].

BRAILLE-LAB AS AN AUTHORIZED APPLIANCE

B-L is an appliance authorized for use by the Hungarian National Federation of the Blind. By March 1987, a total of 95 sets have been installed in the schools of the Federation and by individual users. Based on the speaking BASIC of B-L, the Federation organized two beginners' courses on computation in spring 1986 and 1987. The speaking computer effectively helped the blind participants to acquire knowledge and skill in computation and to put them to creative use. The Users' Manual for B-L has been published on cassette tape and in Braille print as well.

REFERENCES

[1] AB FONEMA: OVE III c Speech Synthesizer Manual. Type 21001.

[2] ANDERS, B.: Digitale Sprachsynthese für Low-Cost Anwendungen. Bauelemente der Elektronik 7. 1981, 246-250.

[3] ARATO A.--KISS G.--TAJTHY T.: A MEA 8000 beszédszintetizátor Commodore 64 számítógépen működő fejlesztő rendszere. /The developing system of MEA 8000 speech synthesizer operating on Commodore 64 computer/ Magyar Fonetikai Füzetek (MFF) 15. 1986, 143-154.

[4] ASTHEIMER, A. Sprachsyntese in LPC-Technik, Elektronik 12. 1981, 73-81.

[5] BOLLA K.: Folyamatos beszédszintetizáló rendszer magyar nyelven (VOXON). MFF 10, 118-129.

[6] BRÜCK v. H. E.--TEULING, D. J. A.: Integrated voice synthesizer. Philips, Technical publication 48. Electronic Components and Applications Vol. 4 No. 2, February 1982.

[7] FONAGY I.: Utószó. In.: Laziczius Gy.: Fonetika. Budapest, 1963, 189-206.

[8] FONS, K.--GARGAGLIANO, T.: Articulate Automata: An Overwiew of Voice Synthesis, BYTE Publications Inc. 1981.

[9] KISS G.: Parolsintezo Kun nelimigata vortaro en la spegulo de la hungara lingvo. In.: Perkomputila Tekstoprilaboro, redaktis Koutny I., Budapest, 1985,33-47.

[10] KISS G. A magyar magánhangzók első két formánsának meghatározása szintetizált hangmintákat felhasználó percepciós kísérlet segítségével. /Determining the First Two Formants of Hungarian Vowels by a Perception Test Involving Synthesized Sound Samples/ Nyelvtudományi Közlemények 87/1. 1985, 159-170.

[11] KISS G.--OLASZY G. Interaktív beszédszintetizáló rendszer számítógéppel és OVE III szintetizátorral. /An Interactive Speech Synthesizing System with Computer and OVE III. Synthesizer/ MFF 10. 1982, 21-46.

[12] KISS G.--OLASZY G.: A HUNGAROVOX magyar nyelvű, valós idejű, párbeszédes beszédszintetizáló rendszer. Információ Elektronika 2. 1984, 98-111.

[13] OLASZY G.: A magyar beszéd leggyakoribb hangsorépítő elemeinek szerkezete és szintézise. Nyelvtudományi Értekezések 121. Budapest 1985.

[14] OLASZY G.--PODOLECZ GY.: A SCRIPTOVOX MEA 8000 beszédelőállító rendszer felépítése és hangelemtára. Kép és Hangtechnika 6. 1986, 49-61.

[15] Revised International Manual of Braille Music Notation 1956. World Council for the Welfare of the Blind, Paris.