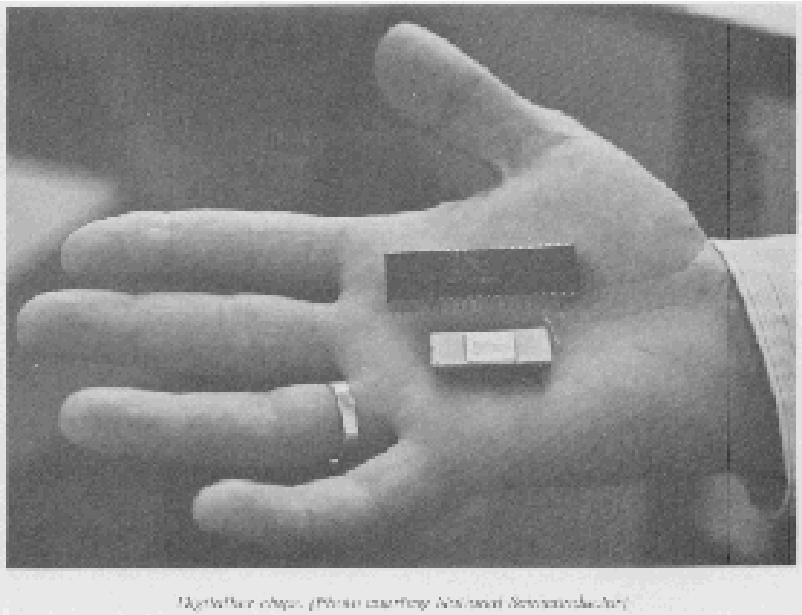
IC Advances in

Advances in Speech Synthesis



Dropbox
Advances in Speech Synthesis

Advances in Speech Synthesis By Larry Nickel
Until recently, speech synthesis was either costly, complicated or cumbersome. The methods of the past were either tape recorders or the use of huge memories to store conventionally digitized speech. Now the technology of large-scale integration (LSI) and very large-scale integration (VLSI), coupled with extensive research into the mechanics of speech, have allowed construction of solid-state speech synthesizers with only a few ICs.

Various Approaches
The tape recorder method was one of the first approaches. It allows storage of many words, but the maximum access time for each word can be very long. The mechanical aspects of this system are not particularly reliable either.
The brute-force approach uses an eight-bit analog-to-digital converter with a clock rate of perhaps eight to ten kHz. The speech data is then stored in ROM (read only memory). For playback a digital-to-analog converter is used. Unfortunately, 100,000 bits per second would require an enormous memory just for ten or 15 words. Instead of using ROM, RAM (random access memory) could be used, with the vocabulary stored on disk and loaded as required. This could impose a great deal of wear on the disk system if the device must speak often.
Some years ago linguists developed a set of sounds called phonemes, which could be strung together to form words. Generally, you have to determine the phonemes needed for the particular words that you want to speak. This can be complicated,

and sometimes the words do not sound exactly as you expect. The advantages are that only a few bytes are required for each word, and virtually any word or sound can be synthesized. Phoneme word synthesizers have been priced at $225 and up. (One supplier is included in reference 1.)

New Product Advances

A number of companies (mostly semiconductor firms) have recently announced voice synthesizer ICs. Usually, the manufacturer provides a ROM with speech data, and this ROM attaches to the synthesizer chip. The speech data contains complex pitch and amplitude information which is processed by the synthesizer. It is not possible for the customer to generate the data that goes into the ROM. The semiconductor manufacturer records the human voice and uses sophisticated algorithms and some human intervention to arrive at the ROM code. The user must either be content with standard vocabularies that are offered or pay huge development costs for custom words.

The good news is that inexpensive standard vocabularies are being developed now, and several will be available at the time of this printing. Industry is looking forward to using this technology for clocks, appliances, test instruments, automobiles, scales, gasoline pumps and telephones. Within the next six months these voice synthesizers will be sold (in huge quantities) for less than $10, and many new electronic products will be talking.

General Instrument

General Instrument (see reference 2) has several interesting products. The SP0256 is a synthesizer, micro-processor and enough ROM for about 16 words, all rolled into one chip. To expand the capability, an SPR16, SPR32 or SPR128 ROM is added. The VSM2032 is a three-chip set intended for a talking calculator or clock and has a fixed vocabulary of 37 word phrases.

Votrax

The same company that start(with phoneme synthesizers years ago is now offering some new product Votrax (see reference 1) has an SC(speech synthesizer IC, an SC(speech pack which says 255 word phrases and an SC01 evaluation (board which includes about 1000 (words/phrases, plus sound effects. Evidently some of this technology included in a new portable product for the handicapped speechless person called the Phonic Mirror Handy Voice.

Texas Instruments

Texas Instruments has two new voice synthesis ICs, using what TI refers to as linear predictive coding. time-varying digital filter models the voice tract, and this filter is excited with a digital representation either glottal air impulses (voiced sounds) or the rush of air (unvoiced sounds).

The bottom line is that getting from voice to ROM code is a very complex process. The TMS5100 is a four-bit device including an on-chip 36 mW power amplifier and is intended for low-cost, high-volume applications. The TMS5200 is an eight-bit device and is for microprocessor and bus oriented uses; it features slight higher quality speech than the 5100 (but requires a low pass filter and amplifier). The 5200 has a READY line and must be treated as a slow memory device by inserting appropriate wait states.

Both the 5100 and 5200 use the Texas Instruments TMS6100 128K-bit ROM which stores the speech data. A demonstration vocabulary is available, but don't buy it because it contains such phrases as Leon thinks it abnormal for a giraffe to roll on the ground. The product of most interest is a standard vocabulary being offered with approximately 200 words and is designated the VM61001.
The custom linear predictive coding speech is very good quality. Additionally, TI is able to generate phonemes. One company, Street Electronics Corporation (see reference 3), is already marketing a phoneme generator based on the TMS5200.
TI also has a technique using allophones. Allophones sound a lot like the Cylon warriors from Battlestar Gallactica-kind of a monotone but understandable. The phoneme or allophone approach allows the speech data to be much more compact.
Before you invest in TI's technology, drop them a note and request literature on the TMS5100, 5200 and 6100 (see reference 4). Also inquire about their speech synthesis chip set, the TMSK201. Evidently the TMSK201 is available with the VM61001 standard vocabulary (but be sure you don't get the demo vocabulary).

National Semiconductor
National Semiconductor has recently introduced Digitalker, their speech synthesis chip, the MM54104. Like the TI TMS5200, the MM54104 requires a ROM for storage. National makes a pair of ROMs designated MM52164-SSR1 and MM52164-SS2, which are packaged with the MM54104 in a product called the DT1050 Standard Vocabulary Kit. The kit costs $85, and the price trend should be down. Regardless, most of the other manufacturers are pricing their products somewhat higher than Digitalker. The ROMs contain 143 letters/numbers and words (see Table 1). I can think of a lot of other words that I would like to have, but these 143 let you do a great deal. Additional standard vocabularies will be available soon. Also, see my software tips below.

Table 1. The Digitalker vocabulary

WORD/PHRASE/LETTER	HEX	DECIMAL
THIS IS DIGITALKER	00	0
ONE	01	1
TWO	02	2
THREE	03	3
FOUR	04	4
FIVE	05	5
SIX	06	6
SEVEN	07	7
EIGHT	08	8
NINE	09	9
TEN	0A	10
ELEVEN	0B	11
TWELVE	0C	12
THIRTEEN	0D	13
FOURTEEN	0E	14
FIFTEEN	0F	15
SIXTEEN	10	16
SEVENTEEN	11	17
EIGHTEEN	12	18
NINETEEN	13	19
TWENTY	14	20
THIRTY	15	21
FORTY	16	22
FIFTY	17	23
SIXTY	18	24
SEVENTY	19	25
EIGHTY	1A	26
NINETY	1B	27
HUNDRED	1C	28
THOUSAND	1D	29
MILLION	1E	30
ZERO	1F	31
A	20	32
B	21	33
C	22	34
D	23	35
E	24	36
F	25	37
G	26	38
H	27	39
I	28	40
J	29	41
K	3A	42
L	3B	43
M	3C	44
N	3D	45
O	3E	46
P	3F	47
Q	40	48
R	41	49
S	42	50
T	43	51
U	44	52
V	45	53
W	46	54
X	47	55
Y	48	56
Z	49	57
AGAIN	5A	58
AMPERE	5B	59
AND	5C	60
AT	5D	61
CANCEL	5E	62
CASE	5F	63
CENT	60	64
400 HERTZ TONE	61	65
80 HERTZ TONE	62	66
20MS SILENCE	63	67
40MS SILENCE	64	68
80MS SILENCE	65	69
160MS SILENCE	66	70
320MS SILENCE	67	71
CENTI	68	72
CHECK	69	73
COMMA	6A	74
CONTROL	6B	75
DANGER	6C	76
DEGREE	6D	77
DOLLAR	6E	78
DOWN	6F	79
EQUAL	70	80
ERROR	71	81
FEET	72	82
FLOW	73	83
FUEL	74	84
GALLON	75	85
GO	76	86
GRAM	77	87
GREAT	78	88
GREATER	79	89
HAVE	7A	90
HIGH	7B	91
HIGHER	7C	92
HOUR	7D	93
IN	7E	94
INCHES	7F	95
IS	80	96
IT	81	97
KILO	82	98
LEFT	83	99
LESS	84	100
LESSER	85	101
LIMIT	86	102
LOW	87	103
LOWER	88	104
MARK	89	105
METER	8A	106
MILE	8B	107
MILLI	8C	108
MINUS	8D	109
MINUTE	8E	110
NEAR	8F	111
NUMBER	9A	112
OF	9B	113
OFF	9C	114
ON	9D	115
OUT	9E	116
OVER	9F	117
PARENTHESIS	A0	118
PERCENT	A1	119
PLEASE	A2	120
PLUS	A3	121
POINT	A4	122
POUND	A5	123
PULSES	A6	124
RATE	A7	125
RE	A8	126
READY	A9	127
RIGHT	AA	128
SS(PLURAL)	AB	129
SECOND	AC	130
SET	AD	131
SPACE	AE	132
SPEED	AF	133
STAR	B0	134
START	B1	135
STOP	B2	136
THAN	B3	137
THE	B4	138
TIME	B5	139
TRY	B6	140
UP	B7	141
VOLT	B8	142
WEIGHT	B9	143

Table 1. The Digitalker vocabulary.

Table 1. The Digitalker vocabulary

Hardware Tips

National doesn't have a fancy name for their digital word encoding technique, but the speech produced is high quality. The hardware interface to a microprocessor is simple. You simply treat the synthesizer just as you would a location in I/O or RAM. Use the CS (chip select) line and the WR (write) lines, which are active low. The MM54104 produces an INTR (interrupt) output when it is finished voicing each word. This can signal your processor to jam in the next word. The CMS (command select) line can be kept low, or, if you would like to reset the INTR line without initiating the next speech word, lift CMS high and perform a WRITE.

If you can easily control the time between words, the INTR line may not be needed. It does help to produce uniform word spacing if INTR is used. To conserve power a ROM enable (ROMEM) pin is provided to shut off the ROM when it is not in use. The speech output from the MM54104 is buffered in a follower (gain of 1) stage. National recommends use of the programmable LM346 OPAMP, probably for low-power applications, but I substituted an LM741. The output of the OPAMP drives an LM386 power amplifier. More gain can be extracted from the LM386 if a 10 uF capacitor (or a cap in series with a resistor) is connected between pins 1 and 8. Be careful of the wiring layout if you do this, because otherwise there could be some oscillation due to the high gain.

I have connected this equipment to my TRS-80 Model I Level II, but it could run on virtually any system. The circuit of Fig. 1 is a good starting point. Fig. 2 shows the TRS-80 pin-outs for the expansion bus at the keyboard and expansion interface. An expansion interface is not required. Be careful how you wire to the TRS-80. The edge connectors match the numbering, for instance, on Viking wire-wrap connector P/N 3VH20/ 1JND5 and many other edge connectors, except for flat cable insulation displacement types. These types are numbered differently.

Note that in the schematic of Fig. 1 the MM54104 is I/O mapped and that an OUT instruction to location 0 will initiate speech and reset INTR. An OUT instruction to location 2 will only reset INTR. In BASIC, for instance, OUT 0,127 would sound the word "ready."

If you leave some extra space on the board when you fabricate this circuitry, you can add some sockets later for additional ROM. With a few TTL chips to manipulate the device selects for the MM52164s, more than two ROMs could be supported. Note that no bus drivers will be needed for the MM54104 regardless of the microprocessor system that it is used with, since there is no way or need to read data from the MM54104. Since the MM54154 is NMOS, it also does not load the bus appreciably. You should write or call National Semiconductor (see reference 5) for data sheets and application notes on the DT1050.

In wiring the DT1050, check and recheck your work with care. A mistake could prematurely destroy the ICs. National recommends using 7-11 V dc to power the MM54104 so the output of the LM78L05 has been adjusted to approximately 9 V dc. I suggest that all power supplies be checked before inserting the synthesizer or ROMs. As with any MOS device, be sure to avoid static discharges.

Software Tips

Any words or letters from the ROM can be chained together to make phrases or sentences. Not as obvious, pieces of words may be combined to make new words. For instance, if you start the word comma and jam in pound halfway through comma, you can get compound. If you gate the audio buffer or amplifier, you could chop off the first part of a word and use the last part. You can use letters as words such as u for you and b for be, etc. I use weight for the word wait.

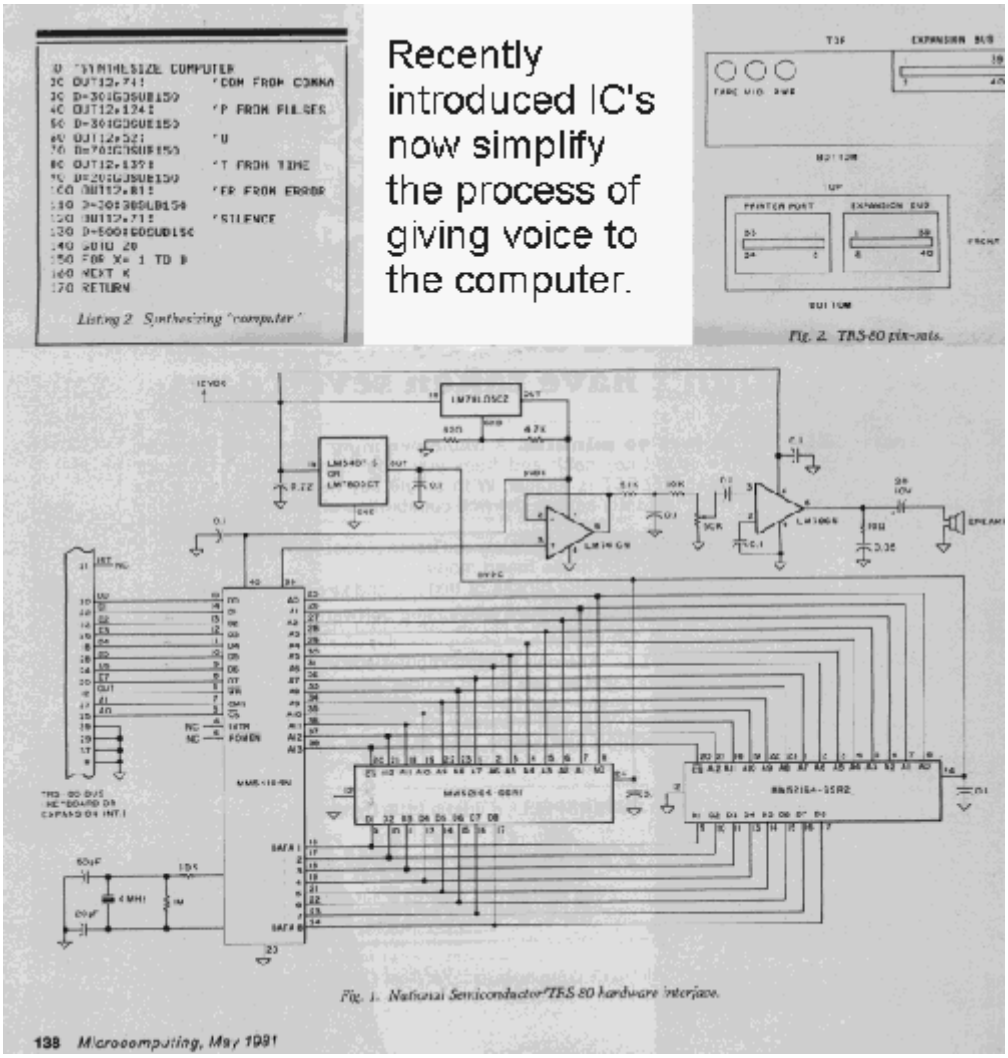
A phoneme-based system would be better for creating your own custom words, but I have found that many new words can be synthesized with only the sounds and syllables in the DT1050. Some short programs are provided as food for thought. Remember that the timing is very critical when only pieces of words are used. If your computer is not a TRS-80, the timing will need to be altered slightly. If you have a TRS-80, then as a starting point type the programs in exactly as they are listed. Note that my Digitalker is I/O mapped as location 12. Change this to suit your system.

The first example (Listing 1) is for the word ATTENTION. The A from again (line 20) is followed by the TEN from the number 10 (line 50), followed by the CH from check (line 80), followed by the A from again (line 110), followed finally by the N

from near (line 130). This gives us A-TEN-CH-A-N. The A-N on the end is used to give an UN sound. The word is understandable but has a slight European accent. Other excamples are COMPUTER, HELLO, YES and NO in Listings 2, 3, 4 and 5, respectively. YES was particularly difficult because the YE sound is not available and a U sound was substituted.
It was certainly not National Semiconductor's intention to use the DT1050 this way, but it does show what can be done with some extra effort. The author would like to know of any other words that the readers may construct.

Listing 1. Synthesizing attention.

```
10 'SYNTHESIZE ATTENTION
20 OUT 12,58
30 D=50
40 GOSUB 180
50 OUTl2,lo
60 D=110
70 GOSUB 180
80 OUTl2,73
90 D=20
100 GOSUB 180
110 OUT 12,58
120 D=1O :GOSUBl8O
130 OUTl2,11l
140 D=25:GOSUBl8O
150 OUT12,71
160 D=500:GOSUBl80
170 GOTO 20
180 FOR X=1 TO D
190 NEXT X
200 RETURN
```



Listing 3. Synthesizing Hello

```
10 'synthesize hello
20 OUT12,90: 'H FROM HAVE
30 D=20:GOSUB110
40 OUT12,43: 'L
50 D=80:GOSUB110
60 OUT12, 103: 'LOW
70 D=70:GOSUB110
80 OUT12,71: 'SILENCE
90 D=500:GOSUB110
100 GOTO 10
110 FOR X=1TOD
120 NEXT X
130 RETURN
```

Listing 4. Synthesizing Yes

```
10 'synthesize yes
20 OUT12,52: 'U
30 D=25:GOSUB90
40 OUT12,81: 'E FROM ERROR
50 D=25:GOSUB90
60 OUT12, 129: 'SS
70 D=500:GOSUB90
80 GOTO 20
90 FOR X=1TOD
```

```
100 NEXT X
110 RETURN
```

Listing 5. Synthesizing No

```
10 'synthesize no
20 OUT12,112: 'N FROM NUMBER
30 D=30:GOSUB90
40 OUT12,46: 'O
50 D=60:GOSUB90
60 OUT12, 71: 'SILENCE
70 D=500:GOSUB90
80 GOTO 10
90 FOR X=1TOD
100 NEXT X
110 RETURN
```

Conclusion

The possibilities for home use of this new technology are many. Computer voice response is especially good where users have nontechnical backgrounds. Hams will now find it economical to provide speech synthesis for automatic station identification or for sophisticated repeater control applications.

References:

1. Votrax Division, Federal Screw Works, 500 Stephenson Highway, Troy, MI 48084 (313-588-2050)
2. General Instrument Corporation, 600 W. John St., Hicksville, NY 11802 (516-733-3107)
3. Street Electronics Corporation, 3152 E. La Palma Ave., Anaheim, CA 92806 (714-632-9950)
4. Texas Instruments, Inc., PO Box 6448, Midland, TX 79701 (915-685-6500)
5. National Semiconductor Corporation, 2900 Semiconductor Drive, Santa Clara, CA 95051 (408-737-5000)

>

Larry Nickel (211 Sacred Heart Lane, Reisterstown, MD 21136) has published articles in 73 Magazine, Electronics World, Ham Radio and other magazines.

source:Microcomputing May 1981 page 134-138