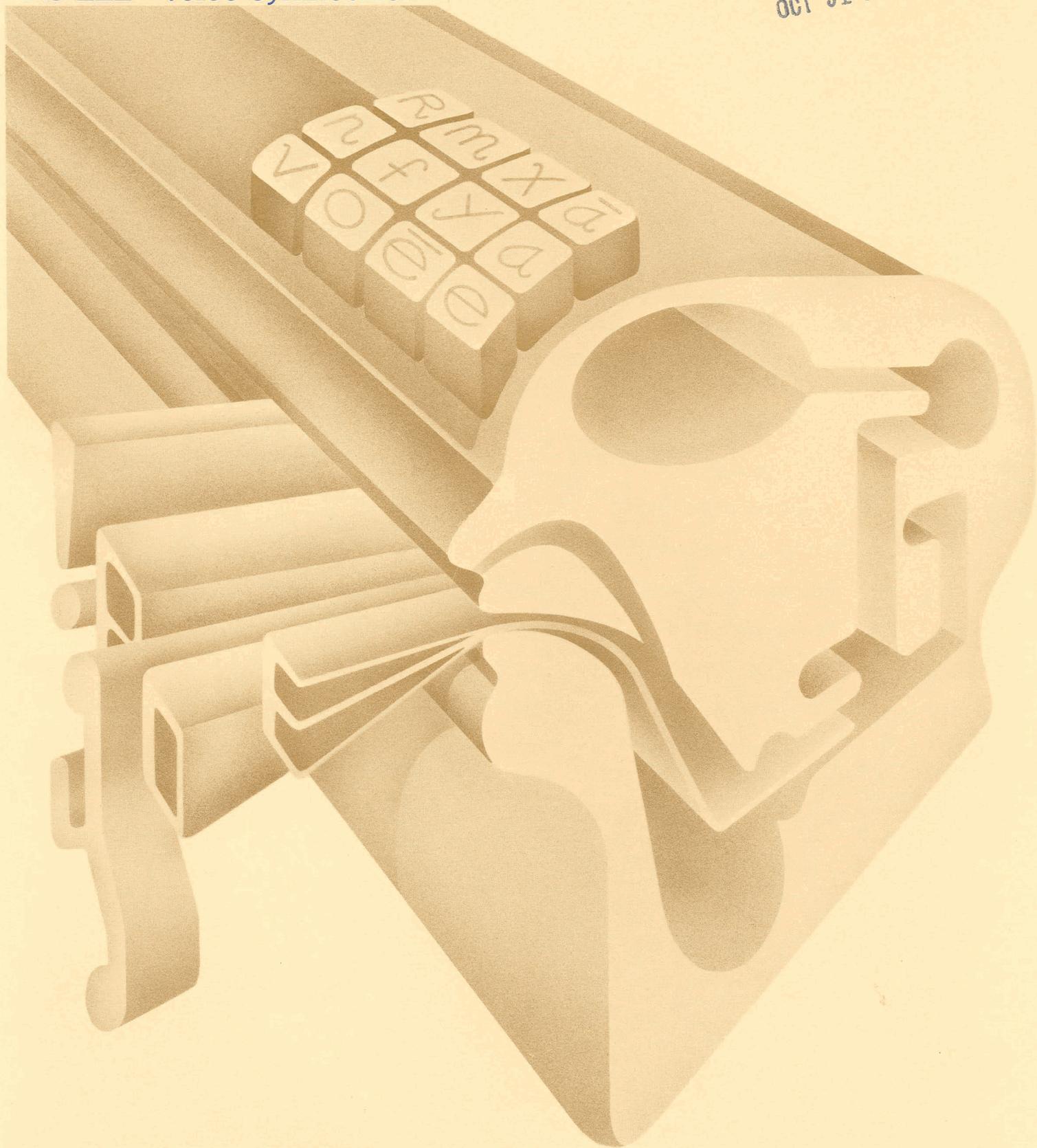


VS III Voice Synthesizer

OCT 31 1972



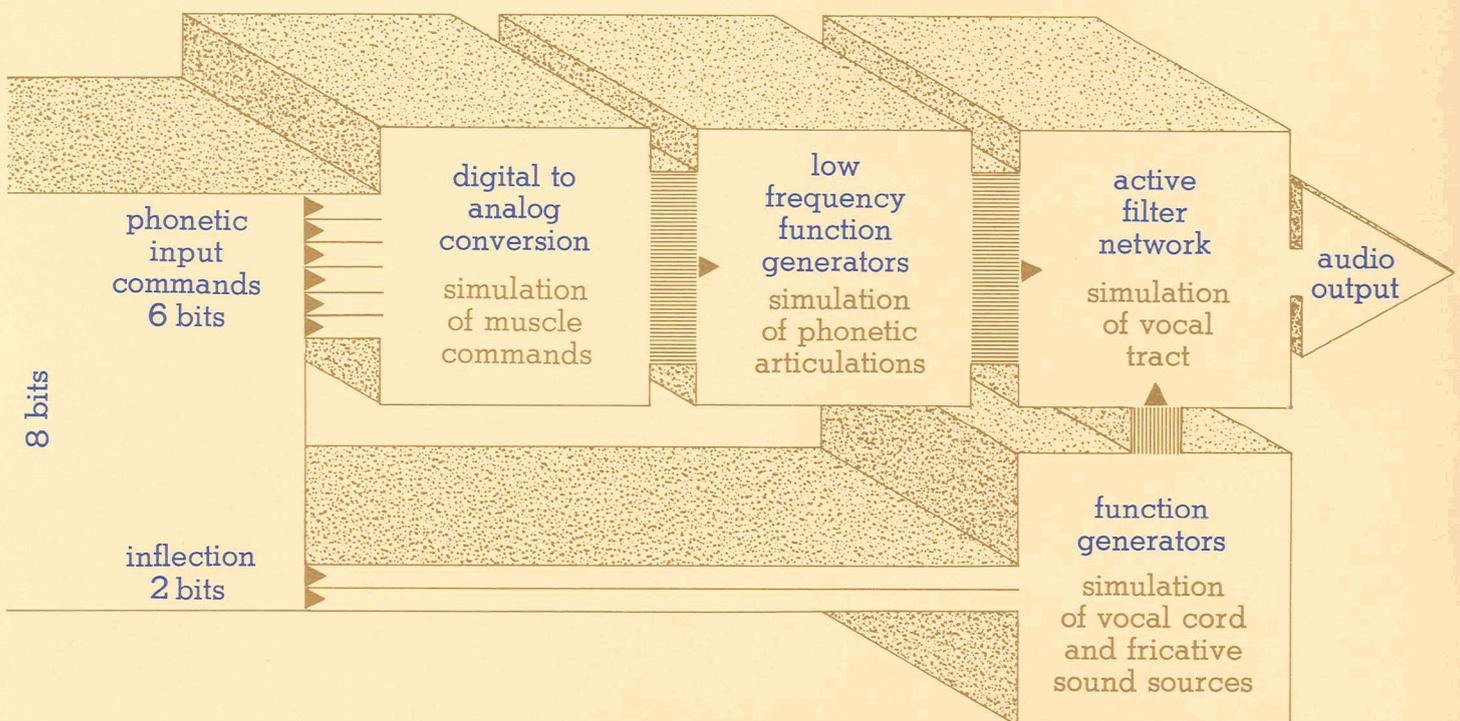
The VSIII Voice Synthesizer is a small, inexpensive system capable of synthesizing human speech with an *unlimited* vocabulary. The unit consists of three circuit cards. It accepts sequences of digital phoneme commands (a phoneme is a basic unit of speech) and converts these commands into corresponding phonetic audio signals. All spoken words consist of these basic building block units called phonemes. The human brain and articulatory system are capable of uttering phonemes in nearly any desired sequence. The VSIII Voice Synthesizer is an electronic simulation of the human brain-vocal system. Thus, any word in the language may be synthesized by a corresponding appropriate sequence of phoneme commands.

Each phoneme command consists of a parallel 8-bit binary word. Typically, there are roughly as many phonemes in a word as there are letters. For example, the word *hello* written phonetically becomes heluō. There are five phonemes in this word, each phoneme command requiring 8 binary bits. Thus, the word "hello" requires 40 command bits of input information to the synthesizer. Six of the eight bits, in each individual phoneme command determine which one of some 50 phonemes is uttered; the other two bits determine the inflection of the synthesized voice for each phoneme. The average word contains six phonemes, requiring six times eight or 48 digital commands. Thus, a 75 word vocabulary will require the storage of 48 x 75 or 3600 bits.

A 128 word vocabulary would require 6144 bits. It is economical to store such small vocabularies in read-only-memory for many applications. Thus, the addition of one more circuit card to its system, containing read-only-memory integrated circuits, will give the VSIII synthesizer a built-in vocabulary of up to 2000 words. The VSIII Voice Synthesizer utilizes analog circuitry to simulate electronically the human vocal system. A short series of appropriate 8 bit digital commands based on the phonetic alphabet will call forth any word in the English language. A series of active filters synthesizes the resonances in the vocal and nasal tracts. Special forcing function generators provide the filter excitation, synthesizing the vocal cords and fricative sound sources. The articulations or dynamics of

the vocal tract parameters when transitioning from one phonetic sound to another are generated by specially designed low frequency function generators. All parameters for proper pronunciation and articulation of each phoneme are specified by a read-only-memory matrix.

The flexibility of the VSIII Voice Synthesizer allows it to operate with its own self-contained vocabulary, or with a vocabulary stored in an external memory, retrieved and transmitted by a computer or other instrumentation. The cost of the unit, with or without built-in vocabulary, will be in an area to justify large volume uses such as in vehicle and instrument communications.



General Specifications

Physical Characteristics

Size 10" x 4" x 11"

Weight . . . 3 lbs. (approximate)

Construction

Solid State Circuitry, Basic three-card system (Modular Design)

Environmental Conditions

Temperature Range 0° - 70° C (wider range available)

Electrical Characteristics

Voltage 117V, 60 cycles A.C.

Power Consumption . . . 10 watts

Audio Output Level

0-10V peak to peak

Frequency Range . . 100 Hz to 5000 Hz

Vocabulary

Size no limit

Word Length . . variable; standard buffer length, 128
phoneme commands

Choice of words and phrases to meet individual requirements

Library

Basic Library stored in ROM . . . 100-500 words

Advanced without ROM unlimited

Speech Components

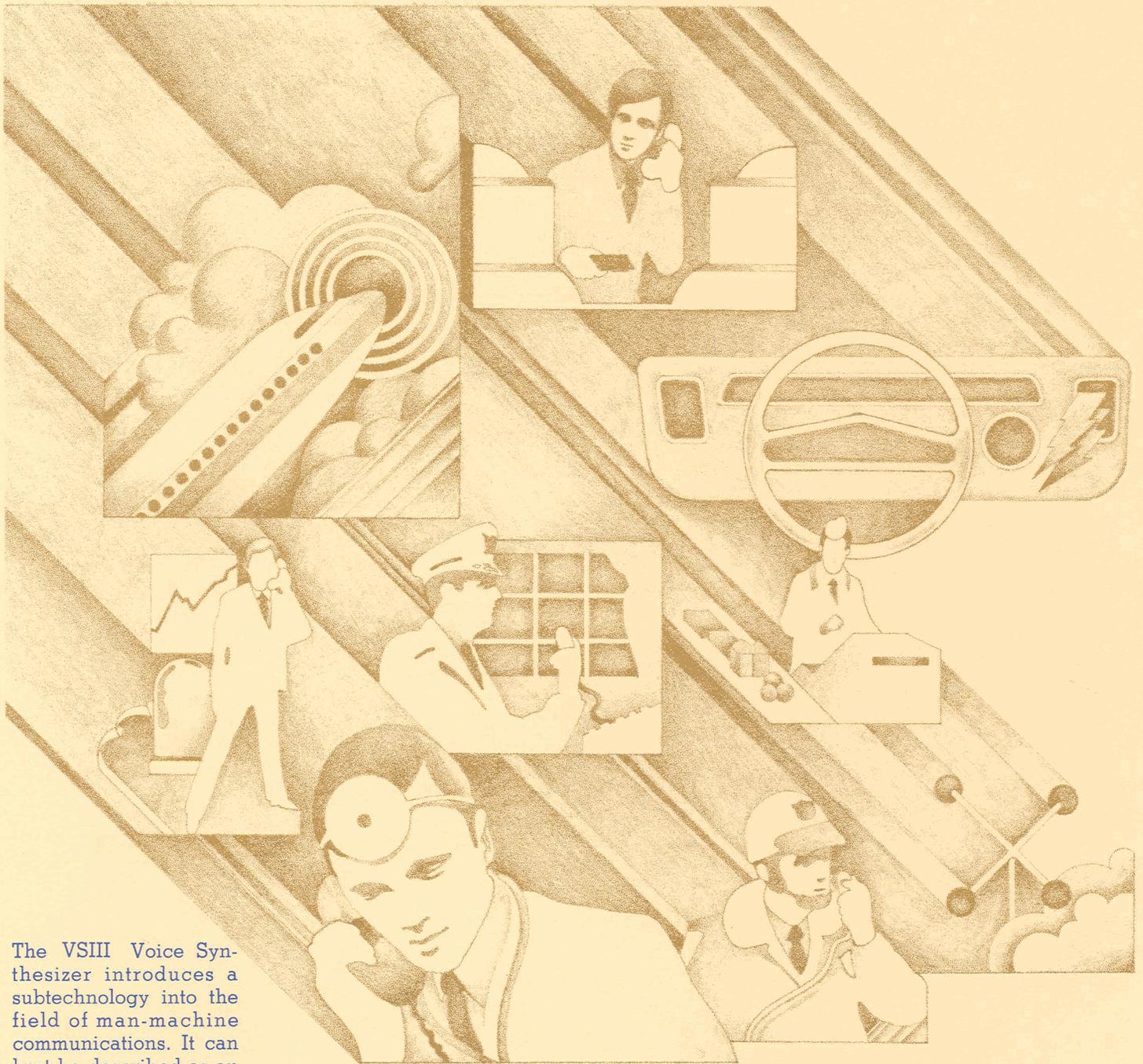
64 phoneme alphabet, with inflection

Voice Generation Method

8 bit parallel phonetic command

Flexibility

Vocabulary and access lines can be changed and upgraded as required by customer.



The VSIII Voice Synthesizer introduces a subtechnology into the field of man-machine communications. It can best be described as an electronic analog of the human vocal system. Its audio output can be transmitted over standard speaker systems or common carrier telephone lines with remarkably intelligible quality.

Several versions of the VSIII Voice Synthesizer are available for use either as a computer peripheral or a stand alone unit. As a peri-

pheral, the device can bridge the man-machine gap by providing direct audio interaction with a computer system. Hence the touch-tone telephone now becomes a responsive computer terminal. As a stand alone unit, with a self-contained memory of variable size (50 to 2000 word vocabulary), the VSIII Voice Synthesizer can readily be integrated

into any system where audio response is desired.

The compact design of this solid state device, and its minimal power requirements, make it well suited for aircraft and surface vehicles. Potential applications for the VSIII Voice Synthesizer cover the full spectrum of Business, Industry and Government.

Vocal Interface Equipment Group

Federal Screw Works

3401 Martin Avenue
 Detroit, Michigan 48210
 Telephone (313) 841-8400 Ext. 316

As appeared in

Electronic News®

New VS-IV Bridges
Man-Machine
Communications Gap

**See preview
of first low-cost,
miniaturized
voice synthesizer**

The VS-IV Voice Synthesizer, no bigger than a household telephone, accepts digital commands from a variety of sources, including any computer, and converts them into completely understandable audio signals. In effect, it permits a computer to speak in understandable English (or, when programmed for the purpose, any other language).

The unit is a miniaturized, solid-state device weighing less than 5

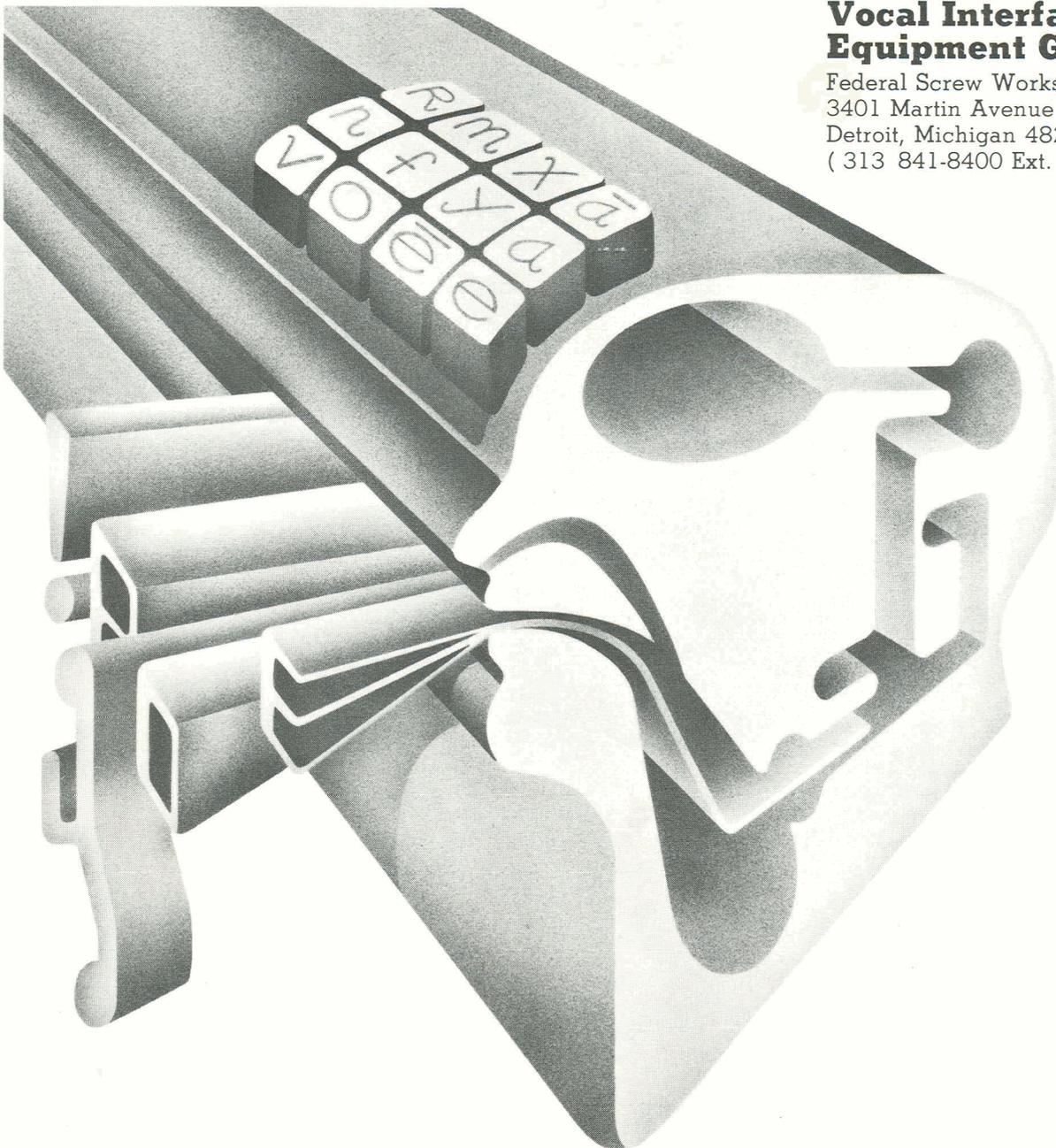
pounds. Essentially, it is an electronic analog of the human voice, receiving the counterpart of brain signals to the larynx, and duplicating human speech through the utterance of the proper phonemes.

The cost of the VS-IV Voice Synthesizer is under \$2,000 in production quantities, positioning it considerably under conventional audio response devices, and suggests its use in applications where other units are impractical.

The VS-IV speaks your language. So do we.

Vocal Interface Equipment Group

Federal Screw Works
3401 Martin Avenue
Detroit, Michigan 48210
(313 841-8400 Ext. 316)



Vocal Interface Equipment Group

3401 Martin Avenue
 Detroit, Michigan 48210
 (313)841-8400 Ext. 316

CLEAN & ADVANCE
ACTIV. KB
OUTPUT
TIED BUFFER
END-TO-START
will be used for ADV
IN/OUT

PROGRAMMING THE VS 4 VOICE SYNTHESIZER

150

PA1	AW	IE	EH	ER	TH	THV	U1	IU	O1	O2	UH	UH1	UH2	IN	OUT
PA2	AE1	AE	W	E	R	T	Y	U	I	O	P	AH	UH3	NOR	REP
PA3	AY	A	S	D	F	G	H	J	K	L	OO	OO1	AH1	ADV	MEM
NG	1	Z	SH	CH	V	B	N	M	I1	I2	IY	2	IN1	IN2	
ZH	A1	IH	3	<i>PA 6 15 ms</i>				4	EH1	EH2	IN3	IN4			

1, 2, 3, 4
IT EASY
BREATHING

Programming the VS 4 is quite simple once the phonetic rules for the language are mastered. Unfortunately, the sounds that are present in a word are not what we think they are, but by following the listed rules, any word in the English language may be synthesized properly.

Many sounds which are considered to be only one basic sound are really made up of a glide of several sounds such as long A, long I, long O, long U, and long E. Other sounds are made up of a pair of sounds such as J and CH.

The language contains basic sounds which receive differing degrees of stress, for instance, the A in the word name and the A in the word namely. The various stresses are designated by a number suffix; the higher the number, the less the stress. For instance:

- UH is full stress.
- UH 1 is stressed less.
- UH 2 is stressed still less.
- UH 3 is very weak stress.

thus, all these commands call forth the same basic sound with varying degrees of stress.

Here is a list of all the phonetic commands and their probable usage.

- AW awful, law
- IE zero
- EH enter, met
- EH 1 seven
- EH 2 seven
- ER weather
- TH three, thick

THV	<u>t</u> his, <u>th</u> en
U	<u>u</u> se, <u>y</u> ou
U 1	<u>u</u> nite
IU	new, <u>y</u> ou
O	<u>o</u> nly, <u>n</u> o
O 1	hello
O 2	no <u>t</u> ice
UH	<u>b</u> t, <u>m</u> ust
UH 1	<u>u</u> ncle
UH 2	stirrup
UH 3	apple (app ___le), able (ab ___le)
AE	<u>c</u> at, <u>s</u> at
AE 1	antenna
W	<u>w</u> on, <u>w</u> ish
E	three
R	<u>r</u> adio, <u>r</u> adar
T	<u>t</u> wo, time
Y	six <u>t</u> y, eight <u>y</u>
I	six, <u>m</u> ix
I 1	<u>i</u> nept, <u>i</u> nside
I 2	cryptic, static
P	<u>p</u> enny, <u>p</u> ound
AH	<u>c</u> ontact, <u>c</u> ar
AH 1	<u>c</u> onnect
A	<u>n</u> ame, <u>c</u> ame
A 1	<u>n</u> amely
S	<u>s</u> even, <u>s</u> ix
D	<u>d</u> o, <u>d</u> iet
F	<u>f</u> our, <u>f</u> ive
G	<u>g</u> et, <u>g</u> rand
H	<u>h</u> ello, <u>h</u> ow
J	<u>j</u> udge, edge
K	<u>c</u> ame, <u>l</u> ock
L	hello, <u>l</u> ight
OO	<u>l</u> ook, <u>b</u> ook
OO 1	<u>g</u> ood, <u>s</u> hould
N 6	<u>r</u> ing, <u>a</u> n <u>g</u> le
Z	<u>z</u> ero, hazy
SH	<u>s</u> how, <u>s</u> hip
CH	<u>c</u> hair, <u>c</u> hime
V	seven, <u>e</u> ven
B	<u>b</u> all, <u>b</u> ed
N	<u>n</u> ine, seven
M	<u>m</u> ile, <u>m</u> en
IY	lie
ZH	azure, pleasure
IH	station, condition

AY may, lay

It is important to know the many glides in English and how to generate them.

Long A in name	A, AY
Long A, unstressed	A1, AY
Long E in three	IE, E or IE, Y
Long I in mile	AH, IY
Long I, unstressed	AH 1, IY
Long O in hello	UH 1, O, here UH 2 may be substituted for UH 1 and O1 or O2 may be substituted for O depending on stress.
Long U in two or new	IU, U
Short A followed by NK like in thank	AE, AY
same short A unstressed	AE 1, AY
OO in book	UH 3, OO
same OO unstressed	UH 3, OO 1
OO in good	IU, OO
same OO unstressed	IU, OO 1
I in is	IE, I
unstressed versions of same	IE, I 1; IE, I 2
OW in cow, OU in cloud	AH, O 1
unstressed version	AH 1, O 1
OY in boy, OI in voice	O, IY
unstressed version	O 1, IY
EA in deaf	IH, EH
unstressed version	IH, EH 1
I in thin	I, IH
unstressed version	I 1, IH, or I 2, IH

IU, IH, UH 2, and UH 3 are frequently used in glides with other vowels. They help produce a more natural conversational rhythm and articulation.

Other multi-part sounds:

J in judge & DG in judge	D, J
CH in catch, match, chair	T, CH
NK in thank	NG, K
NG in engage	N, G
NG in engine	N, D, J

As you can see, many sounds which are normally thought of as single sounds are actually glides which must be represented by two sounds. This characteristic in English is called diphthongization.

We make diphthongs (glides) every chance we get in our speech. This is why a foreigner saying the word three will invariably pronounce it THREEE instead of our usual diphthonged version THRIE. This characteristic of English is not fully appreciated by phoneticians which is one reason why pronunciations of words in dictionaries are usually only of little help in programming the VS 4. With a little practice the user will rapidly learn to use the VS 4 to its fullest capacity, producing well articulated speech, more natural sounding than any other speech synthesis technique to date.

The space bar or PAO produces a short silent interval and is usually used between words to keep the speech from "running together". PA 1 produces a short pause to separate important words. PA 2 produces a medium pause like you would normally find at a comma. PA 3 produces a long pause like you would use between statements. The phoneme commands consist of the binary numbers from 1 to 63. Binary number 0 is treated as a very short silent interval in the synthesizer. This is the natural or rest position for the synthesizer and it will have no output during this command.

An inflection command must accompany each phoneme command. On models with keyboards the appropriate inflection key must be pressed. If no inflection key is depressed, the inflection on that phoneme will be IN 2 which we call the neutral inflection. An advantage of the VS 4 is that under computer control, inflection may be made content-oriented producing a natural inflection impossible to achieve with systems utilizing pre-recorded human speech. This, of course, is not necessary and very intelligible sentences may be constructed from words with preset inflections. Inflection commands on the VS 4 control pitch, timing, and volume just as they are controlled in the human vocal system. In normal speech a stressed syllable is lengthened in time, is louder, and is higher pitched. Inflection commands on unvoiced sounds and pauses are also important because they affect adjacent phonemes. The sequence of inflection commands on a word is referred to as the inflection contour. A good natural contour for short words is 1 2 3 4 or 2 1 2 3. The inflection on longer words depends on which syllables are stressed and which are unstressed, therefore no general rules can be given here.

To enter phoneme commands into the synthesizer merely depress IN key and enter commands as previously described. To output speech, depress OUT key. Be sure to depress NOR (normal) or REP (repeat) keys before OUT key is activated. This will put the synthesizer buffer memory into normal or repeating output mode. The VS 4 will repeat a message endlessly as long as power is supplied.

Vocal Interface Equipment Group

3401 Martin Avenue
 Detroit, Michigan 48210
 (313)841-8400 Ext. 316

PRICE LIST, Effective May 1, 1972

ORDER QUANTITY	VS-IV MODEL A	VS-IV MODEL B
1	\$3,400. EACH	\$3,700. EACH
2	3,120.	3,420.
3	2,880.	3,180.
4	2,680.	2,980.
5	2,520.	2,820.
6	2,400.	2,700.
7	2,320.	2,620.
8	2,260.	2,560.
9	2,220.	2,520.
10	2,200.	2,500.

OEM Quantities Quoted On Request.

MODEL "A": Includes synthesizer, with *no* ROM but with a 128 byte shift register buffer for phoneme input (approximately 15 - 20 words), mounted in a cabinet with power supply.

MODEL "B": Same as Model "A", with an addition of RS-232 interface.

OPTIONS:	Keyboards (Risk Manufacturing) marked with standard phonetic alphabet and/or word list of ROM vocabulary, interfaced to voice synthesizer at	\$875.00 EACH
	Speaker System , including high quality enclosed speaker and amplifier, interfaced to voice synthesizer at	\$150.00 EACH
	Special Portable Cabinet , for mounting of synthesizer, speaker system and keyboard at	\$125.00 EACH

DELIVERY: 60 to 90 days.

WARRANTY: Vocal interface equipment group guarantees each voice synthesizer and related equipment to be free from defective material and workmanship, and agrees to remedy any such defect or to furnish a new or equal part in exchange, if delivered prepaid to Scitronix Corporation, 1600 N. Woodward, Birmingham, Michigan. This warranty covers 120 days labor and parts from the date of delivery. This warranty does not apply to the product if subjected to damage in transit, accident, misuse, or abuse.

FIRST CLASS
PERMIT 25163
DETROIT, MICH.

BUSINESS REPLY MAIL

NO POSTAGE STAMP NECESSARY IF MAILED IN THE U.S.A.

POSTAGE WILL BE PAID BY



Vocal Interface Equipment Group

3401 Martin Avenue

Detroit, Michigan 48210



I WOULD BE INTERESTED IN AUDIO RESPONSE FOR:

APPLICATION(S)

- ON SITE FILE INQUIRY
 REMOTE FILE INQUIRY
 INSTRUMENTATION
 RADIO COMMUNICATIONS
 OTHER (Explain)

CONFIGURATION(S)

- MAIN FRAME
 MINI-COMPUTER
 AS A STAND ALONE UNIT
 8 LINES OR LESS
 MORE THAN 8 LINES

NAME

TITLE

COMPANY NAME

ADDRESS

CITY

STATE

ZIP

BUSINESS PHONE

PLEASE HAVE REPRESENTATIVE CONTACT

PLEASE SEND ADDITIONAL INFORMATION

Telephone (313) 841-8400