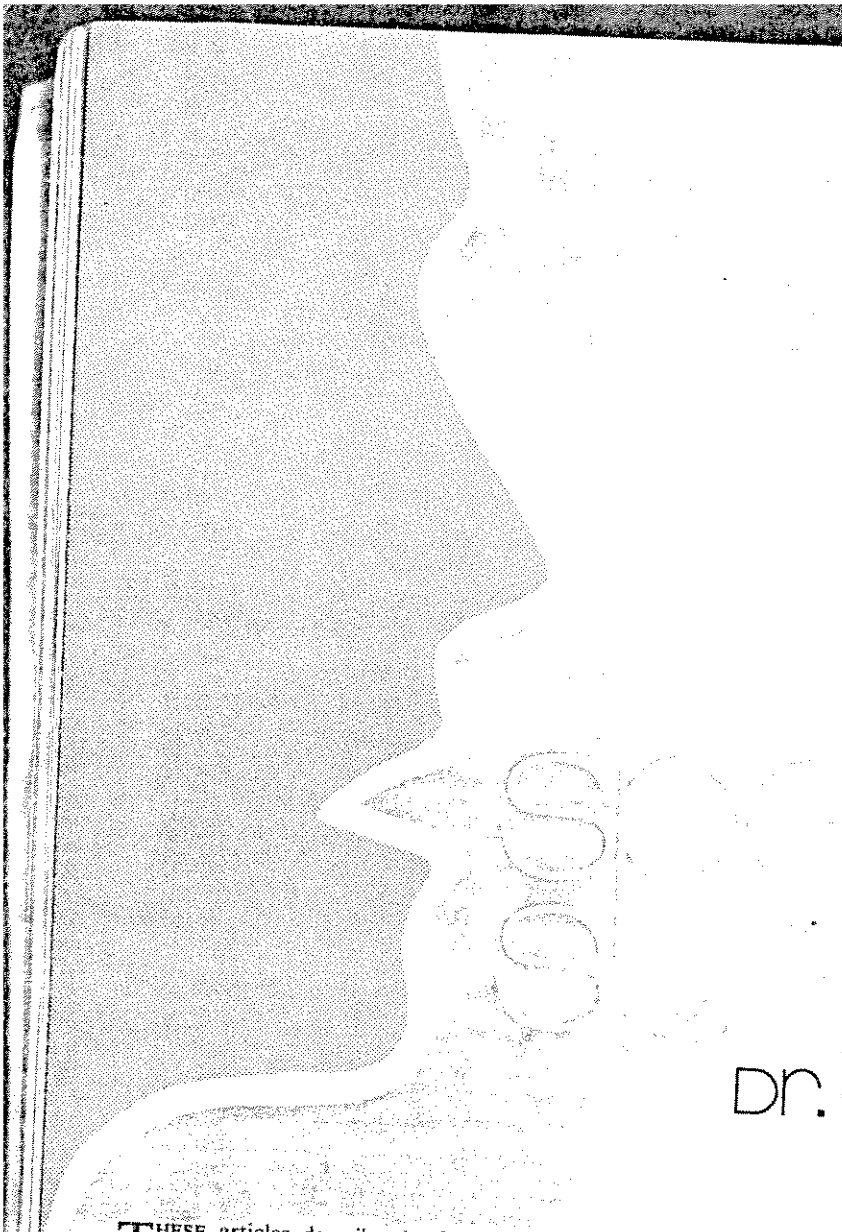


Dr. A.A. Berk · part 1

THESE articles describe the fundamental theory of the synthesis of human speech, and the application of this technique to general microcomputer output. An interface board, developed by Modus System Ltd., and based on an American product, is described as a project. This board, very much as the PE VDU did in its "day", provides any microcomputer user with output in the most up to date manner possible. As examples of its application, interfaces are described for the EDUKIT and the Compukit UK101.

HISTORY

Since time immemorial, mankind has tried to emulate the natural phenomena of bird flight and song, the colours of flowers, human activity and speech, and many other things. To some extent, of course, time has produced a degree of success in each of these areas.

In early Greek and Roman times, apparent emulation of human speech was important for the purpose of convincing people of the validity of such religious institutions as oracular pronouncements and idol worship. Even in those times, technology was of some use in aiding this deception. The priests of antiquity were aware of the properties of the speaking tube, and some instances have come down to us of statues speaking by virtue of such a device connecting an unseen talker with the unmoving lips. In the Middle Ages, speech synthesis was professed by some individuals for less supernatural reasons, namely to prove their ingenuity, and often to impress the monarch. In Pope Sylvester II's reign (999 to 1003), he is said to have constructed a brass head which talked. The immediate result of this was, of course, an accusation of practising magic! In the Thirteenth Century, the very knowledgeable Albertus Magnus constructed an earthenware head which his disciple is said to have smashed in fright upon hearing it talk.

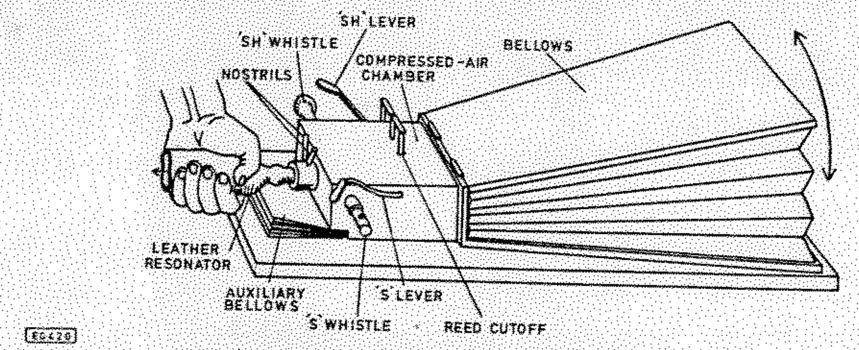
It is possible that some substance may exist for the claims of some of these early workers, by the action of reeds and organ pipes, but the principles offered by some philosophers are amusing at best. It was thought by one that a hollow pipe could be fashioned to collect human sounds and indeed complete sentences, and return them to the ear when uncorked. By this means, he thought, messages could be sent without paper, and talking heads constructed!

Such was the state of the art by the seventeenth and eighteenth centuries: the desire for progress outstripping, as usual, technology's capability for its fulfilment. Deception was still very common, and not very advanced. Charles II and his court were astonished by a speaking head, shown by one Thomas Irson, which answered questions whispered in its ear. At the height of their wonderment, a page discovered in an adjoining room a "Popish priest" answering through a speaking tube. Other deceptions involved the exhibiting of figures seated on boxes filled with bellows and sound boxes etc, which upon very careful examination concealed a dwarf.

In 1779, The Imperial Academy of St Petersburg decided to offer their annual prize for solving the problem of describing the difference between the vowel sounds: A, E, I, O and U, and constructing a device to voice these sounds. Professor C. G. Kratzenstein (the inventor of the vibrating reed resonator as used in the harmonica) solved the problem. He produced five resonant cavities, excited by vibrating reeds, very much as the human vocal chords excite the vocal cavities.

In 1791 came Wolfgang Von Kempelen's invention consisting of a bellows to supply a constant supply of air, as with the lungs, to a vibrating reed and hand operated resonant cavity. Several other cavities and additions such as "hissing" sounds were included to emulate all the human sounds. The device was even capable of uttering complete sentences, a fact which has been later proved by Sir Charles Wheatstone in 1837 by his reconstruction (see Fig. 1) and subsequent improvement to the original device.

The general prevalence of deceptions, however, made it difficult for this first complete mechanical synthesis device to be accepted. Especially as the inventor, Von Kempelen, was himself party to a deception involving a supposed chess playing automaton which won many games before a legless Polish general was found concealed within!



(Scientific American 1972 226:48/58)

Fig. 1. Sir Charles Wheatstone's reconstruction of Van Kempelen's speaking machine. A bellows fed air into a reed which excited the hand controlled resonator. Four separate restricted passages were included for the consonants and nasals.

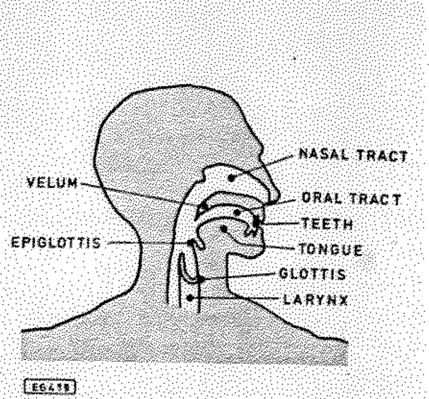


Fig. 2. The Human vocal system consists of two main tracts: The oral tract and the nasal tract. Sounds are produced in the glottis and resonances set up within one or both cavities depending upon the position of the velum.

Von Kempelen's synthesiser marked a major milestone in the art, and remained essentially unmatched until electronics took over, and a machine called the "VODER" was constructed for the New York World's Fair in 1939. Up to this point, the general method of synthesis revolved around copying the actual physics of the human vocal tract, (Fig. 2.)

The physics of natural voice production are those of acoustic filtering. A "wide band noise source" is employed in the glottis, where the vocal chords vibrate under the action of the air forced from the lungs. This vibration excites the cavity between the glottis and the lips, and/or the nasal cavity depending upon the position of the velum. The filtering action which occurs, accentuates certain frequencies, and modifies the sound from the noise source in a complex but well understood manner.

The Voder worked on these principles but took the important step of creating an electronic analogy. Instead of actually expelling air through a set of controllable cavities, a broad band electrical noise source and a random noise generator were filtered electronically, and then amplified and fed to a loudspeaker. The control of the sounds was via a set of hand and foot controls, and as no logic existed for synthesising words, each word had to be produced by the deft manipulation of the controls "in real time". The difficulty of the process, and the complexity of the instrument, can be assessed from the fact that each operator required a year's training, six hours a day, to produce continuous intelligible speech.

PRINCIPLES OF VOICE PRODUCTION

A description of human speech may be given in two different but connected ways. An analysis of the basic phonetic components from which each word in the English language is composed may be performed, and those "phonemes", as they are termed, may be synthesised by a machine, and "called up" as required to produce any given word when written in terms of its phonemes. The other analysis of the voice involves a classification of the exact types of sound which are produced during the enunciation of a word. This would classify the frequencies, speed of production, volume, attack and decay, etc., of these sounds as the word is spoken. To form a word, the necessary frequencies must be produced and mixed in "real time".

The former of these two approaches is termed "phonetic synthesis", and the latter, "frequency synthesis" in the following.

PHONETIC SYNTHESIS

The phonetics of human speech may be split, broadly, into three main components. They are the "voiced", "fricative' and "plosive" sounds. These are described in Table 1. In some cases, two such sounds may blend into one, an example is given under "mixtures" in Table 1. The fricatives are produced by expelling air through a restricted opening such as between the tongue and the teeth. Try saying "this" slowly—you will notice that your tongue starts behind the teeth and moves back. Two fricatives are thus produced one after the other, the first of which is a mixture of a voiced sound (the vowel I) and a fricative (th). By this means, you could build up a complete catalogue of the types of sound necessary to produce each word in the English language. To do this, you would have to classify a few more scales of sound than the three simple ones shown above. This type of analysis is very valuable for speech synthesis, and is called "phoneme analysis". There are some thirty to fifty such phonemes, or basic speech units, from which the majority of English words may be synthesised. If a machine can be built to reproduce these noises, then each word can be written in terms of its phonemes, the information fed digitally to the machine and words produced. Table 2 lists some of the common phonemes with examples of their use, but is by no means complete.

There are several devices on the market which allow you to perform this type of phoneme construction. The word is first written in a code unique to each machine, and the data fed in, often under BASIC. The speech synthesiser then produces the

CLASS	DESCRIPTION	EXAMPLE
VOICED	Continuous sounds such as vowels—the glottis resonates	"a" as in "f<u>a</u>r" "o" as in "t<u>oe</u>"
FRICATIVES	Hissing sounds—the glottis is not required	"s" as in "sound" "sh" as in "wa<u>sh</u>" "h" as in "<u>h</u>at"
PLOSIVES	Short sharp sounds	"t" as in "ha<u>t</u>" "k" as in "ma<u>k</u>e"
MIXTURES	More than one of these sounds may be produced at once	"th" as in "that" is a mix of the fricative "th" and a neutral voiced sound

Human speech may be broken into three main classes of sound or mixtures of these.

TABLE 1

PHONEME		USAGE
VOWELS:	aw	taught
	ae	man
	ah	calf
	a	make
	e	be
	eh	excellent
	er	surge
PLOSIVES:	b	bad
	k	computer
	t	top
FRICATIVES:	th	thanks
	sh	rash
	h	have
SEMI-VOWELS:	w	with
	y	yours
NASALS:	m	man
	n	man
OTHERS:	l	law
	r	ran

English words may be broken down into a number of basic units from which any word may be synthesised—these units are called phonemes. Some of them are illustrated here.

TABLE 2

phonemes thus indicated, in the correct time sequence to be heard as a word. The Microspeech Board by Tim Orr (see *Micro Bus, PE June 1979*) is an example of this type of system.

FREQUENCY SYNTHESIS

To understand the second type of speech synthesis, it is necessary to appreciate the exact frequency structure of the voice sounds themselves. The sound we hear when a word is spoken is composed of several basic frequencies which are mixed together in varying proportions as the word is spoken. In addition, the actual frequency pitches themselves change over the time of the word.

To view the actual frequencies and sounds encountered, the vocal tract should be examined again (Fig 2.) It is clearly seen that there are essentially two resonators available to us during oral speech. These are the full vocal tract from the Glottis to the lips, and the nasal tract. Each of these contributes to word production. It should be borne in mind that they act as resonant filters on the noise produced by the Glottis. The most important cavity of the two, is the tract ending in the lips. In English, the nasal passage is used for only a few sounds (n and m). The main tract may be viewed, ideally, as a single ended resonant cavity of around 17cms in length. (Fig. 3.) Some 'O' level physics reveals that the sort of resonances which such a pipe may suffer are only those of odd-multiple quarter-wavelength as shown. For the length of 0·17m and the speed of sound under normal conditions of 330 m/s, the first three (and most significant) resonances will be as shown. Since the "noise source" which makes the cavity resonate contains all these frequencies, and more, these are available to be accentuated, or filtered out by the cavity. They are called the first three "Formants", and they change continuously during any utterance. Their ranges are given, very approximately. The reason for the frequency variations during word production, is concerned with the complexity of the vocal tract and the constant adjustments made by the musculature during normal speech.

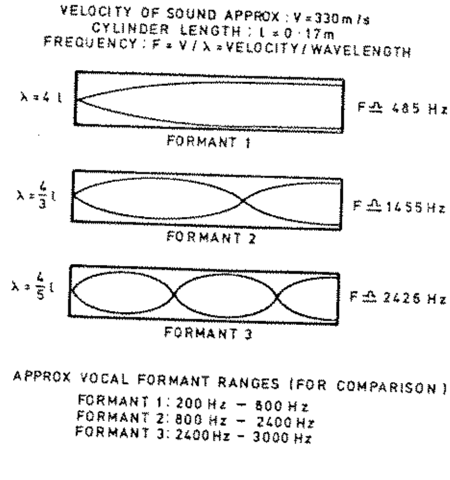


Fig. 3. The first three resonances of a 17cm long cylinder (with one closed end) are shown here. This gives a guide to the main resonances (Formants) of the vocal tract.

These formants may be adjusted in pitch and mixed in different amplitudes to produce the pure voiced vowel sounds. However, they are not sufficient to produce the nasal sounds, which rely on a separate resonator, or the fricatives which are full of random noise. Two more generators are thus necessary for the complete sound. This gives just five different generators to be controlled for a full synthesis. However, these controls involve volume as well as pitch over the time envelope of each word. This type of real time control requires something of the complexity of a microprocessor.

CONSTANT VOWELS

	FI	F2	F3
ee as in seed	240	2300	2900
oo as in wood	420	950	2400
o as in rod	770	1100	2500
aw as in door	570	900	2400

DIPHTHONG VOWELS:

	F1	F2
toy	400–500	1100–2100
prow	400–700	1700– 900

English vowel sounds are produced by adjusting the mix of formants present in the final sound. Diphthongs are produced by smoothly running vowels into each other.

TABLE 3

To illustrate the formant variation during voice production, Table 3 gives typical values to the formants during the constant and diphthong vowel sounds. To appreciate the difference, try saying these sounds to yourself in a normal tone, and listen for the pitch changes, you should not drop your voice as you say the words. The constant vowels use, essentially, constant pitches for the formants, while the diphthong gains its variation by changing the formant pitches as the vowel is sounded. Diphthongs may be considered as mixtures of vowel sounds, which run into each other. The total structure of the voice sounds includes fricative pitch and amplitude changes, as well as nasals. A com-

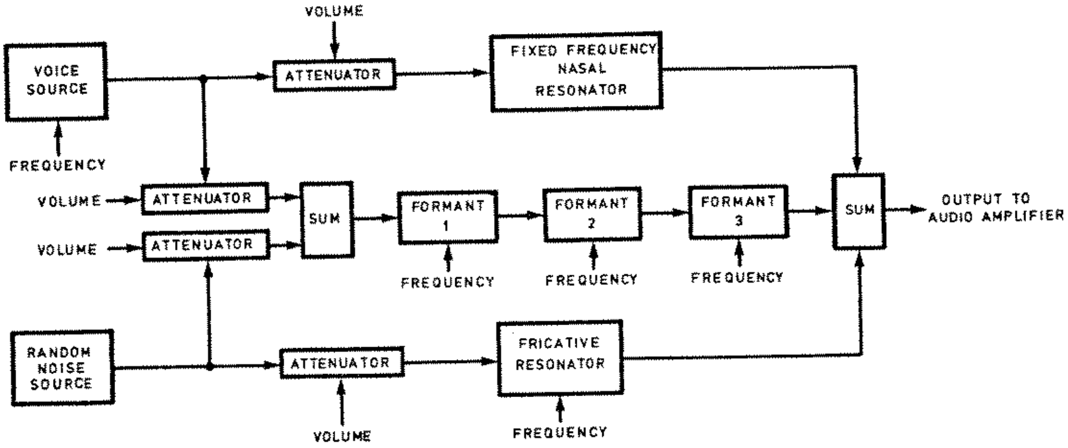


Fig. 4. This "classical" voice synthesis unit requires just nine parameters to specify a single sound (four "volumes" and five frequencies). The "frame" of nine parameters is updated (typically) fifty times a second during the production of a word.

plete list is beyond our present scope, but a fuller picture is introduced below, when electronic voice synthesis is described in more detail.

STORAGE OF SPEECH

One interesting aspect of the above is concerned with the storage of speech in ways other than the conventional manner of grooved disc or magnetic tape. One's first instinct is to think of a system which simply digitises the sound waveforms, and stores them in memory for later retrieval and faithful reproduction. The simplest calculation, based on the evidence, shows that around 8K to 10K Bytes of memory are required for each second of speech-quality recording. This amount of memory is probably best handled by such devices as magnetic tape memories!

The best method for storing speech for output from a computer, is to use one of the approaches suggested by the above two analyses. Words may be stored as phonemes in comparatively small memory blocks, and presented to a "phoneme player" for output. Alternatively, information of frequency, amplitude, fricative content etc is stored digitally, for feeding to a set of sound generators in the correct time sequence, very much as the operators of the VODER would have done.

By one of these means, the common elements of English words are used to reduce the amount of storage necessary for each utterance. The complexity is confined to hardware frequency synthesisers, and controllers.

ELECTRONICS

The electronics of speech synthesis are very interesting, and some of the techniques can now be examined.

Once again it would be tempting to think of some very straightforward approach to the frequency synthesis in the second approach above. Perhaps variable oscillators would be useful, and when their outputs are mixed a perfect synthesis would result. However, it turns out to be considerably easier to attain a natural sound by simply copying nature, and applying controllable resonant filters to the output of a broad band noise generator. (Fig 4.) This provides an almost classical approach to speech synthesis, and would form the electronics for both approaches mentioned above. The first approach would accept phoneme data, and under microprocessor control adjust the frequencies and amplitudes accordingly. The second approach would store the actual values of all those variables and present them to the synthesiser at some sample rate to produce the complete word.

The different sounds arise as follows. The general pitch of the utterance is controlled by adjusting the frequency of the voice source. Pure vowels are produced by one volume and three frequency adjustments to the three formants, and perhaps the nasal resonator volume. Fricatives are produced by a frequency parameter and a volume, and plosives include sudden adjustment to the volume of random noise fed into the voiced channel. In Tim Orr's system, this "frame" of nine parameters is updated fifty times a second. The total effect of the system is considerably greater efficiency of storage than simply digitising speech without analysis.

The other common electronic system is that of Linear Predictive Coding, described below.

LINEAR PREDICTIVE CODING

This technique takes frequency synthesis to a logical ending, and filters, in an "intelligent' manner, the necessary formants and frequencies from a single excitation. It is important to realise that the central principle to which all these techniques are aimed is that of reducing speech to a small set of parameters for efficient memory storage. To this end, the number of parameters needed to specify a sound should be of the order of 10 or 12, and this "frame" should be updated at a rate of around 40 or 50 times per second. The number of bits per parameter in a frame is the main variable, and the more bits, the more accurately the frame is specified. The Texas Instruments "Speak and Spell" unit is based around a Linear Predictive Coding (LPC) system. This system uses a maximum of 48 bits per frame at 50 frames per second i.e. 2400 bits of storage per second of speech.

Words are coded into special binary patterns by speaking them into a microphone connected to a large computer. On this machine, a program is running to convert the sound into codes for storage, and eventually for driving the synthesis unit. In this manner any word may be stored on ROM for later selection and playing.

Each word has a unique set of addresses in the final ROM, and a custom vocabulary may thus be formed. When a word is selected the bits associated with it are fed to the synthesiser which consists of a noise generator, and a 10 stage digital lattice filter. (Fig. 5.) The bits from the ROM are used to set up the filter components in order to produce a binary number which is fed to a digital to analogue converter. As the word is produced, the binary number changes, and the analogue output sounds like speech. In order to prevent sudden changes between frames, a certain amount of linear interpolation is included in the process.

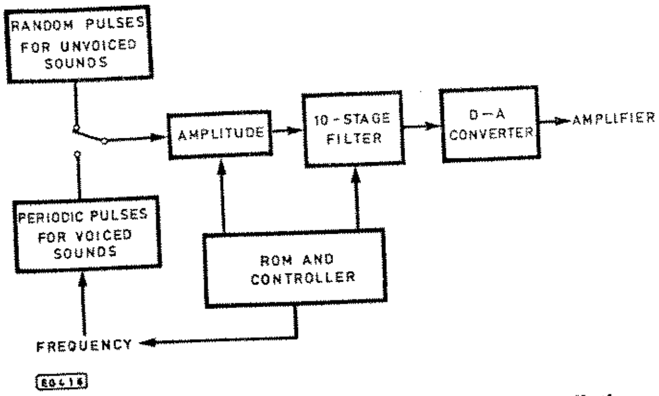


Fig. 5. The Texas Instruments Linear Predictive Coding System codes words into the coefficients of a ten-stage digital filter. A maximum of 2400 bits of data per second of speech are required to adjust the filter coefficients, type of excitation and pitch and frequency as a word is produced.

which is controlled by a form of TI's TMS1000.

The excitation has two time varying parameters, amplitude and pitch. For voiced sounds, a periodic excitation is used, for unvoiced sounds, pseudo random noise of constant amplitude is fed in.

The filter components themselves are quite interesting and along with the controller, ROM and all the other components, illustrates the complexity possible with modern microelectronics. Each of the ten stages (Fig. 6.) contains two adders, two multipliers and a shift register delay circuit. The filtering is thus produced by binary calculation in real time to model the time

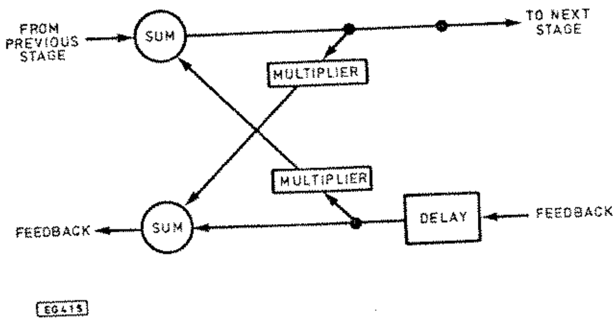


Fig. 6. This illustrates one of the ten stages in the lattice filter used to electronically model the Human Vocal Tract—The multiplication factor (used in both multipliers) is fed from the speech storage ROM.

varying functions of normal speech. A 10 kHz sampling rate is used for the filter; thus during the 100 micro seconds of one sample, two 14 bit additions, and two multiplications are performed. It is due to the difference in speed between these two operations that the delay is necessary. The ROM supplies just the multiplier coefficients for each stage. With judicious use of data, an average data rate of 1500 bits per second of speech may be achieved, giving 165 words, approximately, in a 131K bit ROM. The output sounds excellent, and may be improved upon by incorporating more stages to the filter.

NEXT MONTH: SPEECH SYNTHESISER DESIGN

This project which will add speech synthesis to any microcomputer will be described along with interface examples for two popular microcomputers.

There will also be a special offer of a 24-word speech board available for £39·95 (plus VAT).

and revolve around numerical words such as "two", "thirty", "pounds" etc.

One of the advantages of the fixed-word approach is the extreme ease of interface with any logic system. To aid in the evaluation and addition of speech to any computer or logic device, the interface is described below, as a constructional article. The speech board is available separately, however, with a full technical specification for the user to design his own interface in any way he wishes.

TSI SPEECH SYNTHESIS BOARD

Telesensory Systems Inc (TSI) are the manufacturers of several products for the synthesis of human speech. They produce a talking calculator which has found obvious application for the blind. The units of particular interest, produced by TSI, are the fixed 24 and 64 English word units, consisting of a controller chip (called the CRC), and one or two 2K ROMs, respectively. The photograph shows the chips mounted on a p.c.b. with one or two other discrete components, and a gold-flashed edge connector. The size of the board is 66×73×13mm, with 0·156in. pitch edge connector tracks. The 64-word versions have two ROMs on the board and are slightly larger.

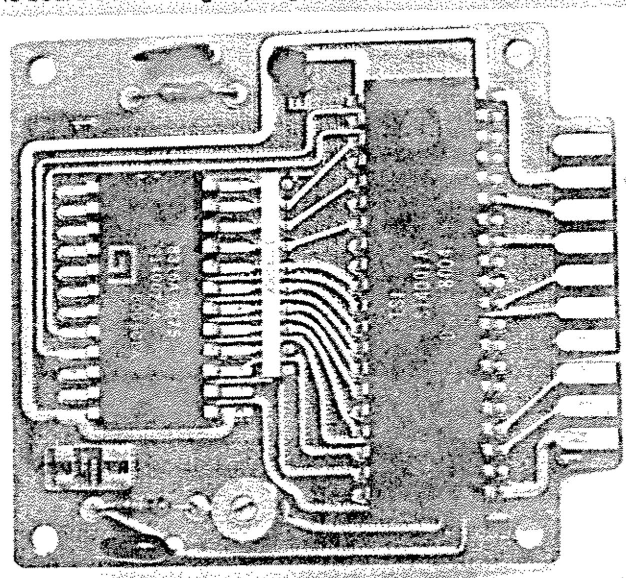
speech synthesis

Dr A. A. Berk · part 2

LAST month, an introduction to the theory and history of speech synthesis was given as a precursor to this month's constructional article. An American speech board (manufactured by Telesensory Systems Inc.) is available through Modus Systems Ltd., and its interface to any general computer or logic system is described. The Compukit and Edukit are used to give examples of the interface, and hence provide readers with a practical demonstration of the way in which speech may be added to their own type of system.

INTRODUCTION

Converting the theory of last month's article into practice at the "one-off" small computer level, was, until recently, rather difficult. Texas Instruments, for instance, have not made their speech products available on this experimentation level, and most other companies are only interested in the large volume "OEM" buyer, who needs expensive customised speech sets. To experiment with the technology, the alternatives include some excellent, but relatively expensive phoneme analysis boards, whereby any word may be synthesised by sending it information of the phonemes which make it up. To "try out" speech synthesis, a far better approach is to use a cheap, mass produced fixed-word synthesiser giving digitally controllable speech on demand from any computer or logic system. If more than one speech set is also available then so much the better. It is with this in mind that Modus have imported a range of fixed-word speech synthesisers, for applications where a customised vocabulary is unnecessary. The words available are basic



TSI speech board

The CRC is a general speech synthesis chip which requires data of pronunciation, pitch, word-length etc. from a ROM to create the necessary sounds. To store the raw data of a word on a read only memory, a complex computer program accepts speech through a microphone and converts it to data suitable to the CRC. TSI have commissioned ROMs, by this method, for several different speech "fonts", and standard vocabularies in a number of different languages (German, French, Arabic etc). The words available here are detailed in Table 1.

S2A	S2B	S2C
(Calculator type	(Standard English	(ASCII 64-words)
24-words)	64-words)	space, X-point, quote,
oh	Same as S2A plus:	number, dollars,
one	ten, eleven, etc.	percent, and, apostrophe,
two	twenty, thirty, etc.	left parent, right parent,
three	hundred, thousand,	star, plus, comma,
four	zero, and, seconds,	minus, point, slash,
five	degrees, dollars,	zero, one, etc. nine,
six	cents, pounds,	colon, semicolon, less than,
seven	ounces, total,	equals, greater than,
eight	please, feet,	mark, at, A, B, C, (etc.) . . . Z
nine	metres, centimetres	lower case, tone,
times-minus	volts, ohms, amps,	upper case, up arrow,
equals	hertz, d.c., a.c.,	control
percent	down, up, go,	
low	stop, low and high tones	
over		
root		
em(M)		
times		
point		
overflow		
minus		
plus		
clear		
swap		

N.B. The ASCII set is of course arranged in exact ASCII order to convert any standard binary code into a verbalisation of the corresponding ASCII character.

TABLE 1. Vocabulary of the three TSI speech boards available

The speech board acts by accepting a control word of six bits to identify the spoken word to be output, followed by a start pulse to tell it to begin speaking. The board is also equipped with a "busy" line to tell external devices that a word is in the process of being spoken. Fig. 2.1 shows a block diagram with timing signals to illustrate the process. The timing shows that the speech output itself does not occur until the "start" signal has fallen to zero. In a typical interface set-up, a set of six latches would feed the control word to the CRC, the "start" signal would fall after 140 micro seconds or more, and the output would be forthcoming. It must be noted that if the "start" line should rise at any time during the speech output, the output is stopped in mid flow. When the "start" line falls again, the word is started from the beginning. To assist in preventing this occurrence, the "busy" line should be monitored by the external system, and the "start" line only operated when the machine has finished talking.

INTERFACING

As is clear from the above, the logic interface requirements are straightforward, and the only problems arise from the rather strange power supply requirements, which are −5V and −15V. However, by a simple trick, the −5V level can be derived from the normal supply of a microcomputer, while the −15V supply must be derived separately.

The final requirement of the board is for audio circuitry of essentially two types, a filter and an amplifier. The sound is produced by the CRC via a digital to analogue converter

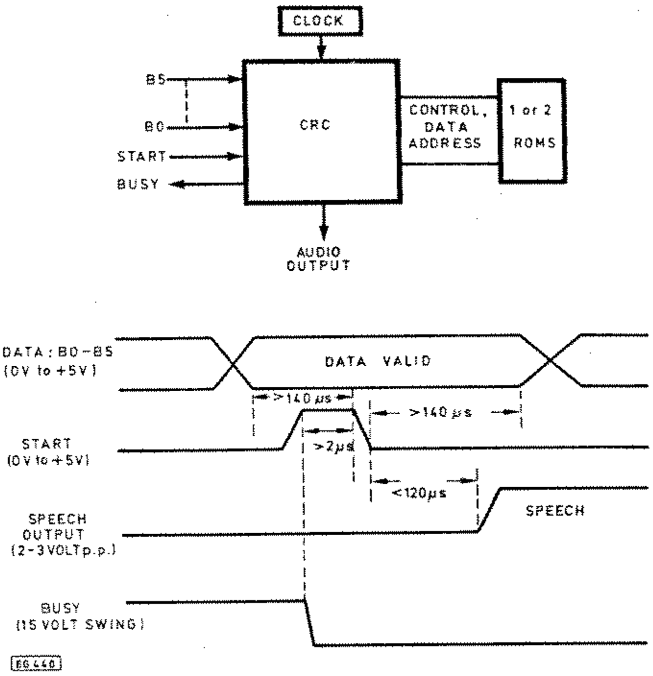


Fig. 2.1. TSI Speech Board

(DAC). This implies that the output waveform is not completely smooth. Its variation from level to level is by a set of steps. A sinewave derived from a DAC will have the form shown in Fig. 2.2. Note that the horizontal width of each step is constant, while the vertical steps vary to approximate the sinewave. This shows that the waveform thus formed may be viewed as a sinewave modulated by a constant frequency squarewave. However, a squarewave contains the elements of many frequencies mixed together (depending on how square it is), and hence the voice output must be passed through a band-pass filter allowing just the band of frequencies needed for voice to pass. To give an idea of the frequencies needed and some typical characteristics, Fig. 2.3 reproduces TSI's suggested op-amp filter response. The three decibel level is often used as a method of comparing such responses, and the frequencies at which this level cuts the graph are called the "corner frequencies". The reason for their importance is that at these frequencies, the incoming waveform's amplitude is exactly halved by passing through the filter.

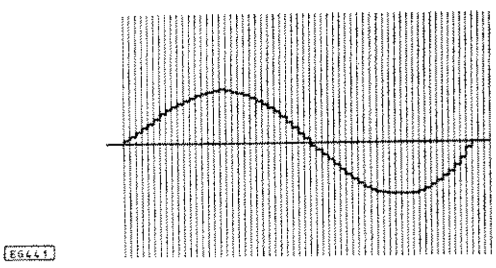


Fig. 2.2. Sine wave output from D to A converter

In considering the TSI output, however, it was found most advantageous to be able to vary the response to different ears, and for different loudspeaker arrangements. After much experimentation, a very simple single op-amp band-pass filter was selected, which, together with an integrated circuit power amplifier gave acceptable results. This circuit is

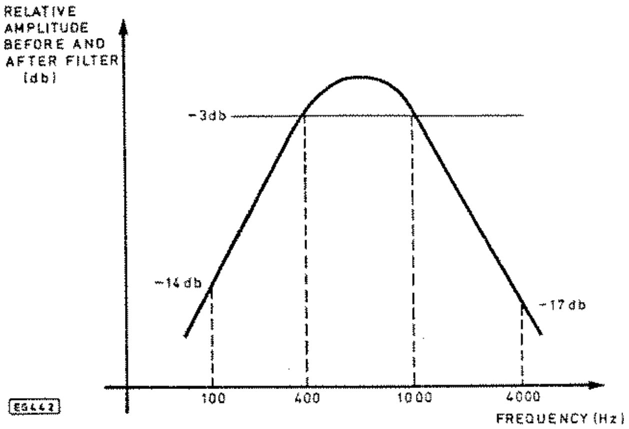


Fig. 2.3. Bandpass filter characteristics

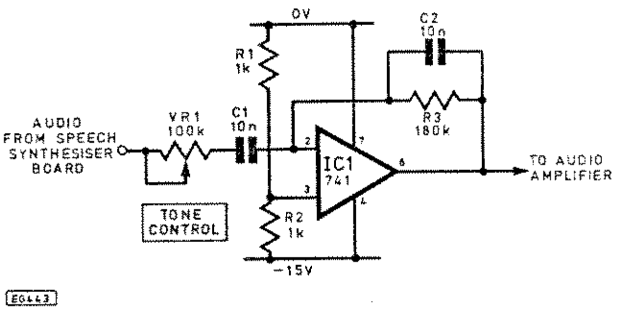


Fig. 2.4. Bandpass filter

shown in Fig. 2.4, and includes a tone control, which varies the filter's pass characteristics.

THE INTERFACE UNIT

As explained above, interfacing is dependent upon three main systems: logic, power and audio sections. All these are included on the interface board described here, along with such important details as the correct edge-connector into which the TSI speech unit plugs.

Fig. 2.5 shows a complete circuit diagram of the interface unit, which is designed to be as complete as possible. A separate 5-volt supply is needed for the logic, as well as a small 10 to 20 volt transformer (at around 250mA) and an 8 ohm speaker.

The word to be spoken is requested by supplying a set of bits (D0 to D5) along with a positive going "latch enable" pulse. The word is latched into IC2 and presented to the speech synthesis board. When the "start" line changes from high to low, and held there, an audio signal is output from the synthesiser. IC1 filters this audio signal in a manner partly determined by the setting of VR1. The filtered signal is a.c. coupled to a volume control and the i.c. power amplifier IC3 which feeds an 8 ohm speaker through C4. The power supply on the board adds 10 volts to the 5 volt supply from the external logic, and just requires an external mains transformer as indicated. The "busy" signal output swings through 15 volts and to convert this to a 5 volt swing, D5 and R5 are included.

CONSTRUCTION

The p.c.b. design and component layout for the unit is shown in Fig. 2.6.

Assembly of the unit is straightforward, and the sockets for the i.c.s should be inserted and soldered first, followed by the edge-connector and ribbon cable connector. Note that the ribbon cable connector has some spare pins for expansion. Discrete components and regulator should be fitted next, observing the correct polarities with great care. It is a very good idea to check that there are 10 volts between −5V and −15V lines from IC4, before proceeding. The 5 volt

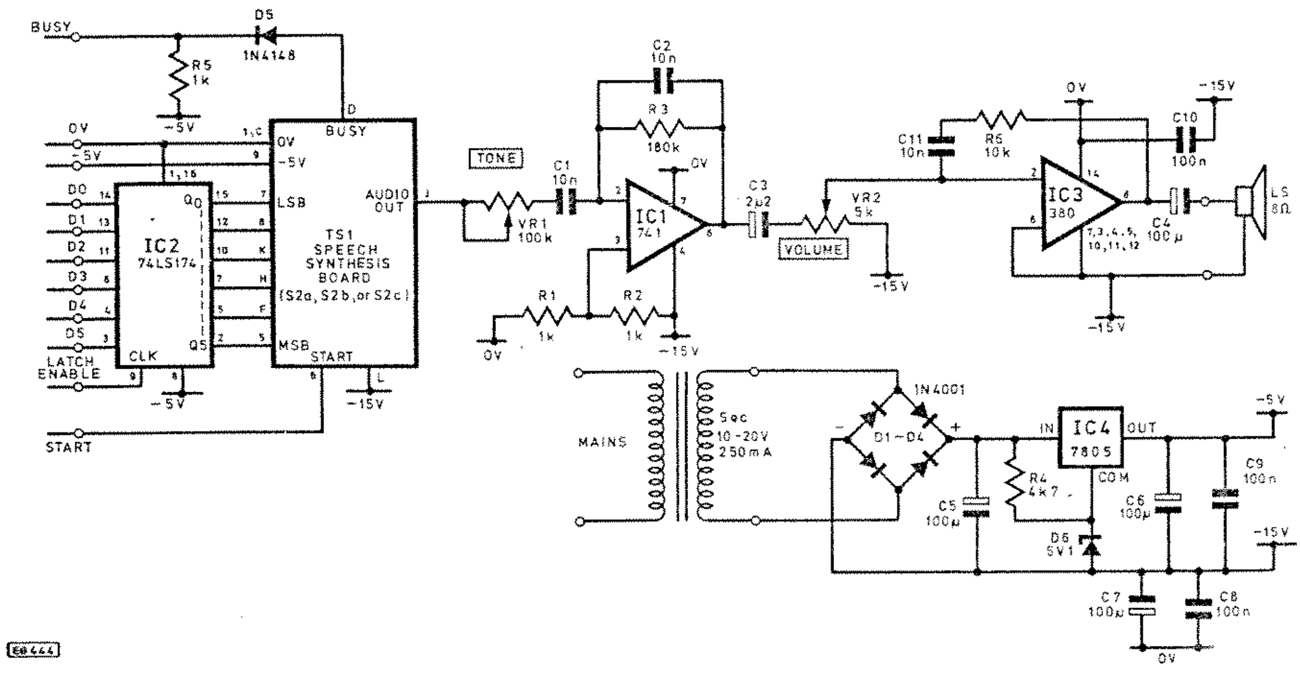


Fig. 2.5. Interface and speech board circuit diagram

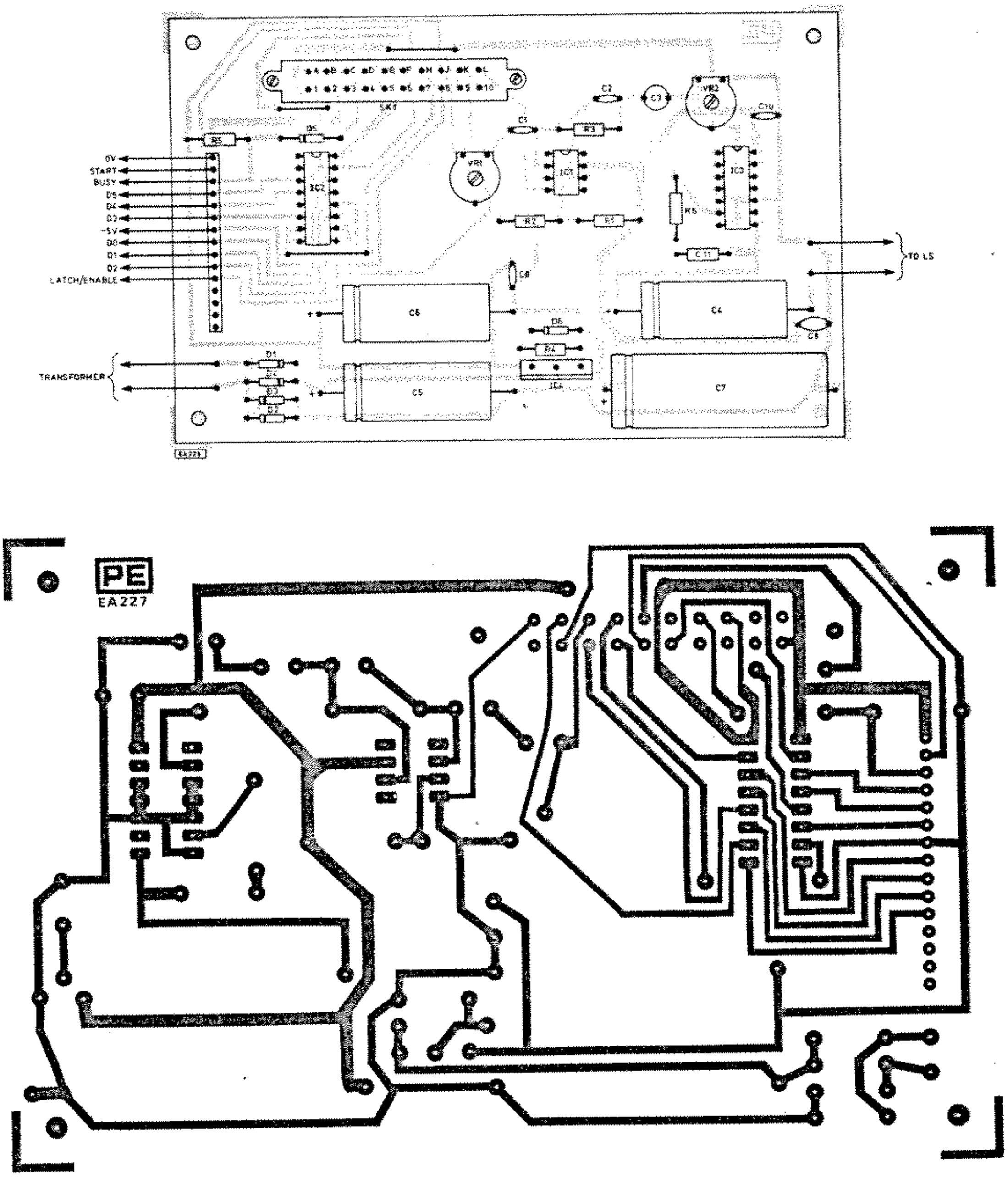


Fig. 2.6. P.c.b. design and component layout for the interface board (*Copyright Modus Systems Ltd*)

supply can then be connected and a check performed for 15 volts across IC1 and IC3 sockets. Check, also, that the correct supply levels are present at the speech synthesis connector and IC2 socket. When you are fully satisfied, switch off and insert IC3. With the volume fully up, a hum should be audible at the loudspeaker when the wiper of VR2 is touched with a finger. Turn the volume down to ensure that this causes the hum to disappear. This will verify IC3's operation.

Next, insert IC1, IC2 (the correct way round) and the speech synthesis board, with the component side inwards as shown. The unit will sound highly unstable unless the "start" line is held at "one" or "zero" (TTL levels). The line should be taken to zero (−5V) through a low resistance (100 ohm for instance). The line may then be taken high and back to zero to check for word output. If none appears, connect all the inputs of IC2 to zero, and pulse the "latch enable" low, then high. This should load the word "Oh". VR1 and VR2 should be adjusted for the best sound. It is useful to note that the word's pitch may be adjusted by the pot on the speech synthesiser p.c.b.—the pot should be set at half way initially—the total number of turns for its full travel is about four and a half, the pot should be turned one way until obviously at its end, and a couple of turns added in the opposite direction. If no sound appears, use a logic probe, meter (or l.e.d. in series with a 1k resistor), to check that the data bits from IC2 to the speech synthesiser board are all zero. If nothing is forthcoming, power should be removed and the whole unit

COMPONENTS . . .

Resistors

R1, R2, R5	1k (3 off)
R3	180k
R4	4k7

All resistors $\frac{1}{4}$W 5% carbon

Potentiometers

VR1	100k hor. preset
VR2	5k hor. preset

Capacitors

C1, C2	10n ceramic disc (2 off)
C3	2µ2 15V elect.
C4, C5, C6, C7	100µ 15V elect. (4 off)
C8, C9, C10	100n disc ceramic (3 off)

Semiconductors

D1–D4	1N4001 (4 off)
D5	1N4148
D6	5V1 Zener BZY88
IC1	741
IC2	74LS174
IC3	LM380
IC4	7805

Miscellaneous

P.c.b.
PL1 ribbon cable plug and connector
SK1 edge connector socket (10 double way at 0·156in. pitch with polarising plug)
Mains transformer 10–20V sec 250mA
8Ω loudspeaker
S2A, S2B or S2C speech synthesis board
i.c. sockets

Constructor's Note

All components including the speech boards are available from **Modus Systems Ltd., 29A Eastcheap, Letchworth, Herts, SG6 3DA (04626 74468/76392).**

The interface board (ex. transformer and loudspeaker) is £14.95 ex. VAT and p&p, and the S2A board is £39.95 ex. VAT and p&p.

checked with the greatest care for any incorrect soldering or component location.

When the unit is working, adjust the two pots on the interface, and the pot on the synthesis board for the best sound. Any tendency toward instability will almost certainly be removed by adding to the power supply decoupling capacitors. Experimentation with the loudspeaker mounting is very worthwhile. Try mounting the speaker in a closed cardboard box with just the speaker cone exposed through a hole of the same diameter. The prototype gave excellent results with a speaker in a transistor radio housing. The speaker should be around two inches or more in diameter for the best results, any smaller and the important base responses will be lost.

COMPUTER SPEECH

It is expected that a number of readers will want to attach the unit to a microcomputer and this section deals with its interface to the Compukit, and the Edukit. The interface requirements for the two machines are sufficiently different to provide an excellent illustration of the process.

In order to operate the speech board effectively, the host computer should have two output lines and one input, as well as the six data lines, usually derived from the data bus, to supply details of the exact word to be spoken. One of the output lines is used to clock the six bit word into the data latch, the other output is used as a "start" signal to the speech board, and the input line is connected to the "busy" line and monitored by the computer to ensure that a new word is not requested before the current one is finished. This last line is not necessary if a delay loop is used by the computer to ensure that the next word is never output too soon.

The Edukit has several input and output lines for these purposes, and its interface is very straightforward. Fig. 2.7 shows a tried and tested arrangement. The Edukit is based around the RCA 1802 MPU which has a number of special and useful features for hardware control. One of its lines, called the Q line, is a flip-flop output whose condition may be set by a couple of dedicated machine-code instructions—one to set and one to reset. This output is used here as a "start" signal. The Edukit has a transistor connected to the Q line, and a couple of pins on the board give access to its collector (through a 100 ohm resistor). A load resistor (2k2 at least) should be connected across the pins, and the collector end of the resistor (pin nearest to the keyboard) should be connected, as shown, to the "start" line of the interface. Thus, the Q output from the 1802 is inverted, and starting of the speech output is effected by first setting Q (giving zero at the "start" line), resetting and setting Q again. This provides a short pulsed "one" on the "start" line.

The Edukit also has a number of input lines, called "External Flags" (EF lines). The states of these lines may be examined by a comprehensive set of jump instructions in the 1802. The "busy" line, therefore, may be examined by EF1, for instance, and the next word sent when the "busy" line returns to a "one".

To transfer data from the computer, the lower six data lines are used, and a further output line is necessary to latch these data bits into the interface's latches. Again, the 1802 is well designed to allow the presence of "output" data on the "data bus" to be signalled by a particular set of output lines (N0 and N2). On the Edukit, one of these lines is inverted and used to allow data to be shown on the digital display. The signal appears at pin 3 of the display drivers, (IC11 or IC12), and it is from here that the latch enable of the interface should be drawn.

To ensure that the unit is interfaced correctly, write a program to turn the Q line off, then on, then off again.

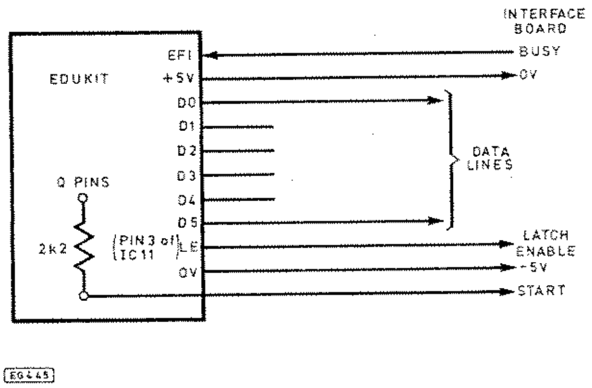


Fig. 2.7. Edukit interface board

Whatever word is held on the latch (IC2) will now be heard when the program is run. If this is working, the hex equivalent of the word spoken will appear on the digital display.

Fig. 2.8 gives a flow chart of a typical piece of software to output all the words offered by the speech board. Either 24 or 64 words are offered on the board, and this will decide when the last word has been spoken. The routine starts by setting the Q line to a "one", and setting the contents of the memory location (called "I" here) to the number 0, which is the first word to be spoken. The OUT instruction is then used exactly as for outputting to the digital display. "I" is then incremented and Q set to 0 for a short time. The word spoken should agree with the digital display reading. A loop is then entered which simply repeats until the "busy" line returns to a 1 level. When this happens, a check is done to determine whether the last word has been spoken; if not the process is repeated, otherwise the end is reached. This routine, does, of course, form the basis for operating any system, though the exact manner in which the checks are made, and the data output, depends upon the hardware set-up involved.

Interfacing to the Compukit is a rather more difficult business as there are no I/O lines on the board. There are two approaches which may be considered. Which is chosen depends upon whether a quick experimental set-up is sufficient, or a fully operational unit capable of controlling the speech fully is required.

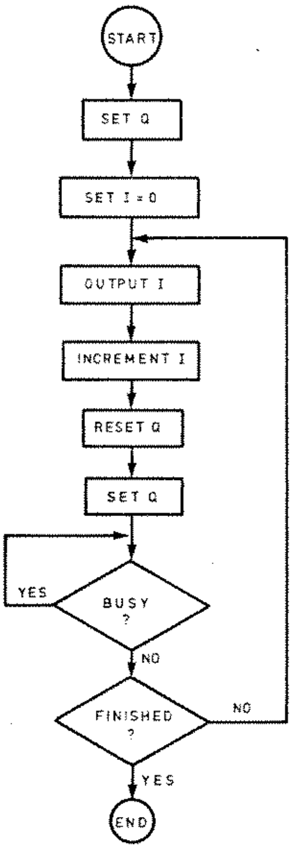


Fig. 2.8. Flow chart to output all the words offered by the speech board

A quick interface to the Compukit, which is by no means a finished set-up, but will give a flavour of the use of speech for a later more sophisticated approach is shown in Fig. 2.9. Here, the top 2K of RAM i.c.s are removed, and their "data" lines, "supply" lines and "chip select" lines are used to drive the speech unit. Notice that the RS lines from the Compukit need inverting before being able to drive the "start" line. The easiest solution is to use any small signal transistor for the job. This interface proved perfectly satisfactory, and should be realised by a couple of d.i.l. headers, rather than by soldering to the p.c.b. When the Compukit is reset and cold started, the Monitor does a memory test, and hence the RS lines are activated. This should cause a word to be output. The theory of operation of this interface is as follows. When a word is to be output, a POKE statement is performed to any memory location in the top 1K of memory. This causes RS7 to go low when the "data bus" contains the desired word. The word is then latched by the speech interface. The next operation is to POKE any location in the next to last 1K of memory—any value will do. This causes RS6 to fall to a zero for a short time, which, via the transistor invertor,

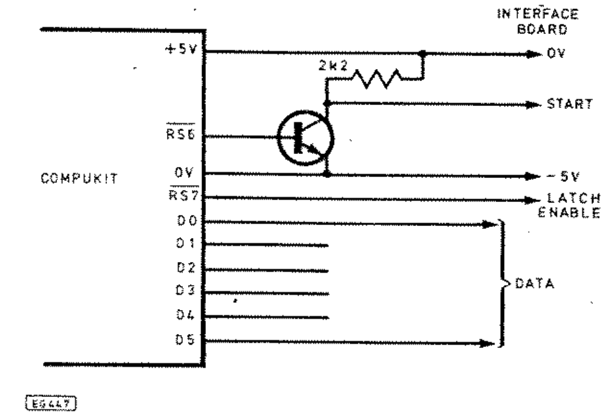


Fig. 2.9. Simple Compukit interface

causes the "start" line to rise for a short time and the speech board outputs the latched word. No provision is made for monitoring the "busy" signal, and if a string of words is to be output, then it is necessary to include a delay loop such as:

FOR I = 1 TO 1500 : NEXT I

between each word output.

The upper 1K of memory starts at decimal 7168, and the next 1K below at 6144. Thus, for instance, the following program will "say" all the words on the 24-word version of the speech board.

```
10 FOR J = 0 TO 23
20 POKE 7168,J
30 POKE 6144,0
40 FOR I = 1 TO 1500
50 NEXT I
60 NEXT J
```

The upper value in line 40 depends upon the frequency of the clock on the speech board, and hence may need to be adjusted to ensure that nothing is lost.

The above interface should not be viewed as a long term method of providing the Compukit with speech. The correct approach is to use a PIA type device such as that detailed in the PE Compukit articles. By this means, the "busy" signal can be used to ensure that nothing is lost, and the top 2K of RAM remains free. An excellent and general interface project for the Compukit is described in a set of articles by D. E. Graham, starting in the next issue of PE. The project is perfectly set up to control the speech board as well as many other things. Another advantage of this interface is that it allows the use of digital to analogue converters. The need for such devices, with reference to speech, is mentioned below.

GENERAL NOTES

Any digital device which displays decimal information, such as a digital clock, digital car instruments, test gear etc. should be able to be interfaced with the speech unit described here. Some demultiplexing may be required, and the on-board latch provides the basis of such circuitry. Though a major application is for computer speech, a computer is by no means necessary to drive the unit.

In using the speech unit, it is worth pointing out that any word spoken to the human ear out of context may be misheard or misunderstood by the listener. It is therefore a good idea to listen through the complete vocabulary, at least once, while following a written copy. This should adequately attune the ear to the sound being produced.

To experiment with voice output fully, the pitch of the utterance must also be controlled in real time. The speech boards allow the clock frequency to be supplied from an external source, and an external D to A converter followed by a voltage controlled oscillator would be a method of allowing computer control of pitch. Volume could also be adjusted at IC3, and, along with pitch, ultimate control of the process would result. The next step would be to produce a piece of software which automatically determines and controls the pitch and volume depending upon context and syntax of the utterance. Although the speech unit presented here is relatively simple, it is more than adequate for these and many other sophisticated experiments in speech synthesis. Indeed, the unit is actually useful, and provides the user with yet another source of output from electronic equipment. This is especially true when the eyes are busy with other tasks, but constant monitoring of numerical data is required.

Finally, the author would like to thank Mr. M. Terkow of Modus Systems for his assistance in prototyping and checking many of the ideas presented in these articles. ★

Readout...

A selection from our Postbag

Readers requiring a reply to any letter must include a stamped addressed envelope.
Opinions expressed in Readout are not necessarily endorsed by the publishers of Practical Electronics.

Velikovsky: Frank Hyde Replies

I offered to answer specific questions from Mr. Austin but so far none have come to hand. However since Messrs Warlow and Williams have added their criticisms on this matter it gives me an opportunity to assume that, certain main themes are themselves statements of belief if not scientific fact, accepted by these gentlemen.

First let me make it clear that I am dealing with the matter from my own original edition of Velikovsky's "Worlds in Collision". Most of the book consists of statements without any support other than quotations from random works of the past. Much of the support for his own ideas are drawn from the biblical text. It is here that quotation is regarded as confirmation of physical facts and as is all too common with this kind of "proof" is not sufficient to substantiate the claims.

Dominant among the statements which serve to antagonise physicists and astronomers are the claims that a comet was ejected from Jupiter. Let us take this as a starting point and I quote page 48 of "Worlds in Collision" *"In the middle of the second millenium before the present era, as I intend to show, the earth underwent one of the greatest catastrophes of its history. A celestial body that only a short while before had become a member of the solar system—a new comet—came very close to earth".* Then for some 14 pages of quotations of various happenings interspersed with statements such as *"the tails of comets are composed mainly of carbon and hydrogen gases. Lacking oxygen, they do not burn in flight, but the inflammable gases, passing through an atmosphere containing oxygen, will be set on fire".* Well the earth passed through the tail of Halley's comet in 1910 but no one noticed it nor was the tail visible, let alone on fire. It is perhaps wise to point out here that the "tail" of a comet appears as the comet approaches the body which determines its perihelion passage. If the comet is large the "tail" will be produced in a direction away from the body it approaches and will continue as it passes round the body to point away from it. It may be that more than one "tail" appears. Sometimes the tail is insignificant and sometimes there is no visible "tail". Halley's comet is due in 1986 and between now and then there is to be a mission to observe the comet at first hand. Thus will all doubt as to the composition be settled. Beyond saying that there is an abundance of literature available about comets which when read will give some of the answers, I will leave the actual comet in order to test the *statement* by Velikovsky that Venus was originally a comet ejected from Jupiter. This he says occurred around the year 1500 BC. The Book of Exodus provides the details of what happened when the comet grazed the earth. Many very strange things are told such as, the houses of the Egyptian people were destroyed, the Nile ran red etc. The curious thing about it all was that the houses of the Israelites were not affected. After forty years the comet came back and caused further trouble, to go off and again return, and, after hitting the earth, again bringing it to a standstill before going on to upset Mars into its present orbit and itself settling down as Venus. The earth regained its place and continued on as it was, before all these events took place. All this supported by historical facts quoted from the Bible and other writings from various parts of the world. These are very far reaching statements and thus need careful examination. Many so-called scientific predictions have gone awry even when evidence appears to be adequate to make a forecast. However there are certain everyday facts of natural science which must be satisfied before new and astounding statements can be accepted. Let us then look at what the statements of Velikovsky imply in relation to comets.

The basis of many of the statements attends on "belief". Because many of the cometary orbits lie near Jupiter a number of people including Laplace and Pierre Simon put forward the view that Jupiter might be the source of cometary bodies. However there are no writings in support of the fact nor indeed are there any sightings of it ever happening nor has Jupiter ever been observed to have unusual events occurring in its vicinity. One person, V. S. Vsekhsviatsky does believe that comets are ejected from volcanoes on the satellites.

I will not resort to writing out the mathematics in this reply but will give the consequences as are known and used in astronomical physics. Of course if anyone would like to have the full mathematics I will send them, if a stamped addressed envelope is supplied. The effect of projection of a portion of Jupiter of the density and size to end up as a planetary body like Venus, would be to raise the temperature of Jupiter several thousands of degrees centigrade. This would have melted the body being ejected and therefore it would probably be dispersed as dust or vapour. Thus it is unlikely that even a comet (assuming that a comet is not anything like current thinking) could exist or survive such an experience. There is another problem also. The escape velocity at Jupiter is of the order of 20km/second. Whatever the escape mechanism might be it would not be aware of this fact for if the velocity of escape was 70km/second the comet/planet would fall back into Jupiter, if it were 73km/second it would escape from the solar system. In either case it is more likely that Jupiter would be considerably changed and not its stable self as has been observed over a far longer period of time than as recently as 1500 BC.

There is still a further problem. This is the mass of Venus. It amounts to rather more than 5×10^{27} grammes. The total kinetic energy that would be required to propel Venus to the escape velocity of Jupiter is of the order of 10^{41} ergs. This poses an even greater problem than all the others put together for 10^{41} ergs is equivalent to all the radiation energy of the sun for a year. Or in other terms one hundred million times more powerful than the largest solar flare ever observed.

A final word on this situation, which is occupying more space than is justified. Velikovsky has quoted several rapidly occurring collisions involving planets yet the odds against it happening once in a millenium is 30,000 years. Surely unwritten folk stories and legends must defer to practical and demonstrable facts, for this is why David Birch was dismissive.

Frank W. Hyde, FSE,PEng, FRAS.

This correspondence is now closed—Ed.