MILLSGRADE

3 Halcyon Court
St Margarets Way
Stukely Meadows
Huntingdon
Cambridge
PE18 6ES

Tel: (0480) 413298

VOXBOX
Speech Synthesis System
User Guide

MODEL NO. VX-1

Copyright (C) 1984
NOTE: For use with the Acorn Electron computer only

ALLOPHONE SPEECH SYNTHESIS AND VOXBOX

There are several ways of synthesising speech electronically, most of which involve the storage of complete words and phrases in memory. These techniques require large amounts of memory as they typically can store between seven and ten words per 8Kbit (1Kbyte) of memory. The user is also restricted in choice to the words in the system's memory and as the memory capability of most home microcomputers is not large enough to provide a reasonable selection of words, the vocabulary of such a system would be very limited. The technique used by VOXBOX provides an unlimited vocabulary and is called allophone speech synthesis.

Like other systems, VOXBOX has a fixed number of sounds stored in memory but there is an important difference - the stored sounds are not complete words. A word can be broken down into a number of basic unit sounds. It is these sounds, called allophones, which form the VOXBOX's library of stored sounds. For example the word "chin" can be broken down into the unit sounds of "ch-", followed by "-i-", followed by "-n". A spoken language has a set of these unit sounds or allophones from which the words of that language are constructed. VOXBOX uses the English language allophone set from which the words of the true English language can be constructed. Table 1 shows the 64 allophones which make up the VOXBOX's allophone set and the sounds they represent. Sample words are given as examples of the type of sounds that make up the VOXBOX's allophone set. It is useful to go through this table pronouncing all the sample words and listening to the allophones within the words.

Along with the obvious advantage of an unlimited vocabulary allophone speech synthesis uses much less memory as only a small number of short duration sounds have to be stored. VOXBOX even stores this data in its own (ROM) memory leaving your Electron's memory available for other uses.

SILENCES		
PA1	10ms	
PA2	30ms	
PA3	50ms	
PA4	100ms	
PA5	200ms	
VOWELS		
AA	100ms	lot, what, chop
AE	120ms	hat, extract, lapse
AO	100ms	water, aught, talk
AR	290ms	arm, cart, garment
AW	370ms	ground, shout, down
AX	70ms	mud, lapel, succeed
AY	260ms	try, mine, light
EH	70ms	enter, pet, heavy
EL	190ms	label, apple, chappel
ER1	160ms	letter, over, furniture
ER2	300ms	fern, burn, learn
EY	280ms	great, late, wait
IH	70ms	pit, ring, loaded
IY	250ms	people, knee, heavy
OR	330ms	poor, boar, law
OW	240ms	loan, grow, close
OY	420ms	boy, noise, choice
UH	120ms	pull, hook, book
UW1	100ms	computer, beauty, dispute
UW2	260ms	zoo, food, two
XR	360ms	chair, care, their
YR	350ms	near, peer, hearing

CONSONANTS		
BB1	80ms	brick, tub, rubber
BB2	50ms	bat, bend, box
CH	190ms	church, creature, teacher
DD1	70ms	end, maid, loaded
DD2	160ms	drag, desk, dot
DH1	290ms	this, then, they
DH2	120ms	bathe, path, with
FF	150ms	fish, cough, physical
GG1	80ms	guest, gate, gear
GG2	50ms	greed, gut, good
GG3	160ms	log, anger, garment
HH1	130ms	hit, hair, heat
HH2	180ms	how, harm, hurt
JH	140ms	jug, injure, fudge
KK1	160ms	crown, kite, care
KK2	190ms	leak, mask, park
KK3	120ms	clown, cook, core
LL	110ms	list, hello, cymbal
MM	180ms	milk, charm, ample
NG	220ms	running, hung, anger
NN1	140ms	win, groan, nut
NN2	190ms	annoy, normal, not
PP	210ms	pot, ample, heap
RR1	170ms	red, wrong, ring
RR2	120ms	bread, crate, train
SH	160ms	ship, wish, action
SS	90ms	sister, moss, dose
TH	180ms	thin, thought, thick
TT1	100ms	its, pests, hats
TT2	140ms	train, tatty, hat
W	190ms	vest, move, ever
WH	200ms	white, when, twig
WW	180ms	wood, we, queen
YY1	130ms	beauty, cute, computer
YY2	180ms	year, yes, yoga
ZH	190ms	beige, treasure, measure
ZZ	210ms	zoo, maze, praise

TABLE 1. VOXBOX's allophone set

USING ALLOPHONES TO CONSTRUCT WORDS

To use allophone synthesis successfully you need to become familiar with the technique of breaking words down into their constituent allophones (basic sounds). This will become easier as you get used to the contents of Table 1. It is necessary to think of the sounds, and not merely the letters, from which a word is constructed as there is no one-to-one relationship between the written letters of a language and the sounds. For example the "oa" sound in the word "load" sounds the same as the "o" sound in the word "mode" and the "ou" sound in the word "though" and would therefore use the same allophone, yet the "gh" spelling in the words, ghost, enough and dough, sound different and would therefore use different allophones.

Say the word "hat" slowly to yourself. Now say it again whilst trying to break it down into as many different sounds as possible. You should find that you can break the word "hat" down into three sounds or allophones, H-A-T. These allophones have the symbols HH, AE and TT. The allophone representation of the word "hat" is therefore "HH, AE, TT".

Say the word "chat" to yourself. Now try and break this word down. Although there are four letters you should have found that it was only possible to break "chat" down into three sounds, CH-A-T. These allophones have the symbols CH, AE and TT.

Say the word "plat" to yourself and break it down into its constituent allophones. Unlike the "ch" spelling "pl" will break down giving four sounds, P-L-A-T with allophone symbols, PP, LL, AE and TT.

The following is a list of words broken down into their allophones and their allophone representations. Reading this list should help you start thinking about words being composed of sounds and not merely letters.

s-i-tt-i-ng	SS,IH,TT,IH,NG	p-i-ll	PP,IH,LL	e-n-d	EE,NN,DD
m-e-n	MM,EH,NN	a-c-t	AE,KK,TT	th-a-n	TH,AE,NN
c-oo-k	KK,UH,KK	p-u-ll	PP,UH,LL	au-ght	AO,TT
w-a-t-er	WW,AO,TT,ER	l-a-p-e-l	LL,AX,PP,EH,LL	s-u-n-k	SS,AX,NN,KK
h-o-t	HH,AA,TT	wh-a-t	WH,AA,TT	wh-ea-t	WH,IY,TT
p-eo-p-le	PP,IT,PP,EL	b-ai-t	BB,EY,TT	g-r-ea-t	GG,RR,EY,TT
l-i-ke	LL,AY,KK	s-k-y	SS,KK,YY	n-oi-se	NN,OY,SS
t-oy	TT,OY	c-u-te	KK,UW,TT	f-oo-d	FF,UW,DD
m-oa-n	MM,OW,NN	c-l-o-se	KK,LL,OW,ZZ	s-ou-n-d	SS,AW,NN,DD
m-ou-se	MM,AW,SS	p-oor	PP,OR	f-or-t-u-ne	FF,OR,TT,UW,NN
a-l-ar-m	AE,LL,AR,MM	l-a-ugh	LL,AR,FF	n-ea-r	NN,YR,ER
ea-r-r-i-ng	YR,ER,RR,IH,NG	h-air	HH,ZR	sh-are	SH,XR
w-i-sh	WH,IH,SH	w-i-g	WW,IH,GG	j-u-gg-le	JH,AX,GG,EL
y-ear-n	YY,ER,NN	v-er-se	W,ER,SS	p-l-ea-s-ure	PP,LL,EH,ZH,ER
b-ei-ge	BB,EY,ZH	q-u-i-ck	KK,WW,IH,KK	s-q-u-are	SS,KK,WW,XR
c-l-ou-d-y	KK,LL,AW,DD,IY	p-r-ai-se	PP,RR,EY,ZZ	b-a-th-i-ng	BB,EY,DH,IH,NG

A LITTLE MORE INFORMATION

The precise sound and duration of an allophone is changed by the position it occupies within the word. It is also affected by the other allophones around it. To cope with this, the VOXBOX's allophone set contains where necessary several versions of the same allophone i.e. there are three KK allophones KK1, KK2 and KK3. The following points are examples of the various ways in which the sound of an allophone may be changed.

Some allophones sound better at the front of a word than at the back. For example the DD1 allophone is used in the word "load" where it occurs at the back of the word, whereas the DD2 allophone is used in the word "data" where it occurs at the beginning of the word. If a DD allophone occurs in the middle of a word, it will either be at the end or at the beginning of a syllable as in the words "head-ing" and "a-dore" where DD1 and DD2 are used respectively.

If you refer back to Table 1, you will notice that the allophone DD1 is of shorter duration than DD2. Initial consonants are generally of greater duration than final consonants. The allophones FF, SS and TH can be doubled when used at the beginning of a word, for example in the word "sister" the first "s" is a doubled allophone and the second "s" is a single, giving an allophone representation "SS,SS,IH,SS,TT2,ER1".

Some allophones (BB, CH, DD, GG, JH, KK, PP and TT) require a pause before them. For most of these VOXBOX has already put in a brief silence but if you decide a longer pause is needed then one of the PA allophones can be used. These pause allophones are also used to give silences between words and clauses.

Some allophones sound better before certain vowel sounds than others. For example the KK1 allophone should be used before the vowel sounds AE, AR, AX, AY, EH, EY, IH, IY, XR and YR whereas the KK3 allophone should be used before the vowel sounds AO, AR, OR, OW, OY, UH and UW.

SILENCE

PA1 (10ms)	- before BB, DD, GG and JH
PA2 (30ms)	- before BB, DD, GG and JH
PA3 (50ms)	- before PP, TT, CH and KK and between words
PA4 (100ms)	- between clauses and sentences
PA5 (200ms)	- between clauses and sentences

VOWELS

AA*	- pottery, cotton
AE*	- extract, acting
AO*	- talking, song
AR	- farm, alarm, garment
AW	- sound, mouse, down
AX*	- lapel, instruct
AY	- kite, sky, mighty
EH*	- extent, gentlemen
EL	- little, angle, gentlemen
ER1	- letter, furniture, interrupt
ER2	- monosyllables: bird, fern, burn
EY	- great, statement, tray
IH*	- sitting, stranded
IY	- treat, people, penny
OR	- fortune, adorn, store
OW	- zone, close, snow
OY	- noise, toy, voice
UH*	- cookie, full
UW1	- after clusters with YY: computer
UW2	- in monosyllabic words: two, food
XR	- hair, declare, stare
YR	- hear, earring, irresponsible

CONSONANTS

BB1	- final position: rib; between vowels: fibber
BB2	- initial position before a vowel: beast
CH	- church, feature
DD1	- final position: played, end
DD2	- initial position: down; clusters: drain
DH1	- word - initial position: this, then, they
DH2	- word - final and between vowels: bathe, bathing
FF*	- fish, cough, physical
GG1	- before vowels: YR, XR, IH, IY, EY, EH
GG2	- before vowels: UW, UH, OW, OY; and before clusters: green, glue
GG3	- before vowels: AE, AW, AY, AR, AA, AO, OR, ER; and middle clusters: anger; and final: peg, dog
HH1	- before vowels: YR, IY, IH, EY, EH, XR, AE
HH2	- before vowels: UW, UH, OW, OY, AO, OR, AR
JH	- judge, injure
KK1	- before vowels: YR, IY, IH, EY, EH, XR, AE, AY, ER, AX; initial clusters: cute, clown, scream
KK2	- final position: speak; and clusters: task
KK3	- before vowels: UW, UH, OW, OY, OR, OA, AR; initial clusters: crane, quick, clown, scream
LL	- like, hello, steel
MM	- milk, alarm, ample
NG	- string, anger
NN1	- before vowels: YR, IY, IH, EY, EH, XR, AE, ER, AX, EH, XR, AE, ER, AX, AW, AY, UW; final clusters: earn
NN2	- before vowels: UH, OW, OY, OR, AR, AA
PP	- pleasure, ample, trip
RR1	- initial position: read, write
RR2	- initial clusters: broad, crane, grease
SH	- shirt, leash, nation
SS*	- sister, moss
TH*	- thin, thought
TT1	- final clusters before SS: tests, its
TT2	- all other positions: test, street
W	- vest, prove, even
WH	- white, whim, twenty
WW	- we, quest
YY1	- clusters: cute, beauty, computer
YY2	- initial position: yes, yarn, year
ZH	- beige, pleasure
ZZ	- zoo, phase

* These allophones can be doubled

TABLE 2. Allophone application information

Some short duration vowel sounds (AA, AE, AO, AX, EH, IH and UH) can be doubled when a longer sound is required as in stressed syllables. For example the second syllable of the word "extend" uses the doubled allophone EH, EH giving an allophone representation "EH,KK1,SS,TT2,EH,EH,NN1,DD2".

Some allophones sound better in allophone clusters (scr.., pl.., br.., etc) than others. For example of the two RR allophones RR2 sounds better in clusters like crown, grass, bread.

Table 2 contains information to help you decide which allophones to use in which words and in which positions within the words. Due to the many subtle variations in speech this table can only give general guidelines as to allophone selection. It is also inevitable that due to these variations some words will sound better than others.

GETTING GOING

Warning: Care should be taken when connecting and disconnecting the VOXBOX. Make sure that the power supply to the computer is switched off.

After checking that the power supply is off, connect the VOXBOX to the expansion connector at the rear of the computer. If the VOXBOX and your computer are on a smooth level surface then the VOXBOX will locate itself and slide easily onto the expansion connector.

Turn on the power supply and if you listen carefully a faint hissing should be heard from the grilling on the top surface of the VOXBOX.

Viewing the rear will reveal a slot in the casing. Through this a small sound output level adjustment screw should be visible. This is factory-set to the halfway position but may be adjusted with the aid of a small screwdriver to suit your requirements.

The VOXBOX is supplied with a software cassette containing two programs. Put this in your tape machine and rewind to the beginning of Side A. Program A should now be loaded with the command CHAIN"SPEAK". A short header page will load and run before the main program. When this runs you will hear the following spoken by the VOXBOX, "Your Speaking Electron Computer" - "By Neil Ellicott" - "Loading", then the main program will load. You may have difficulty understanding all the spoken words at first as it is necessary to become familiar with the VOXBOX's accent.

When the main program is loaded, you will be asked "Load file? Y/N". A file containing a catalogue of example words follows program A. This should be loaded by first pressing the "Y" key and then entering "EXAMPLE" + RETURN in response to the prompt "ENTER CATALOGUE NAME". You may want to load your own catalogue eventually instead of the one supplied. In which case, the tape containing your own files should be inserted into the tape machine and your own catalogue name entered.

MAIN SCREEN

Once the catalogue file has loaded, the "main screen" will appear. The main screen is divided into three sections.

The first section is the "allophone selection section" which contains the VOXBOX's allophone set arranged into three groups - PAUSES, CONSONANTS and VOWELS. There is a flashing cursor, the "allophone cursor", over the BB1 allophone which is used to select the allophones in this section. The allophone cursor is moved using the normal cursor keys.

The second section is the "editing line" where the allophones are placed during word construction. There is a solid non-flashing cursor, the "editing cursor", which is positioned at the beginning of the editing line. The editing cursor is moved with the "<" and ">" keys (unshifted). The editing line extends for six lines on the screen.

The third section is the "data line" where allophone application data is displayed. The beginning of this is marked by a chequered flag.

Try moving the allophone cursor. Note how it moves in four directions and jumps from allophone to allophone. The editing cursor moves in just two directions.

WRITING ALLOPHONE STRINGS

The following keys are used:

Cursor keys - Moves allophone cursor
"COPY" - Copies allophone under allophone cursor to editing line

An allophone string is a number of allophones joined together, i.e. allophone representation of a word or phrase. Try entering the following example allophone string:

KK1,AX,MM,PP,YY1,UW1,TT2,ER1

Using the cursor keys, move the allophone cursor to the first allophone "KK1". Press the "COPY" key.

If you look at the editing line, you will see that the KK! allophone has appeared on it and the editing cursor has moved on one allophone position.

Using the cursor keys and the "COPY" key as above enter the remaining allophones onto the editing line. If you make a mistake then press the "E" key and start again.

SPEAKING ALLOPHONE STRINGS

The following keys are used:

"S" - Speak allophone string on editing line
"SPACE" - Repeat the allophone string
"RETURN" - Return to main screen

You should have the following on the editing line:

KK1,AX,MM,PP,YY1,UW1,TT2,ER1

This allophone string is the allophone representation of the word "computer". To speak this string, press the "S" key. The main screen will disappear and be replaced by a mouth which you should hear and see say "computer" - "space to repeat".

Press the SPACE bar and VOXBOX will repeat the allophone string. Press the RETURN key and the mouth will disappear and the main screen return.

EDITING ALLOPHONE STRINGS

The following keys are used:

"<" (unshifted) - Moves editing cursor left
">" (unshifted) - Moves editing cursor right
"E" - Erases allophone string on editing line
"O" - Restores old allophone string to editing line
"R" - Rubs out allophone under the editing cursor
"I" - Inserts space under the editing cursor

Press the "<" key. The cursor will move one allophone position to the left so that it is over ER1 on the editing line. Move the editing cursor left until it is over MM.

Press the ">" key. The editing cursor will move one allophone position to the right so that it is over PP.

Press the "R" key. The PP allophone which was under the cursor will disappear and the editing cursor will now be over the YY1 allophone.

Press the "I" key. A space will be inserted before the allophone under the editing cursor which will now be over the space. Another allophone may be copied into this space. The editing line may now be cleared.

Press the "E" key. The allophone string is erased.

Press the "O" key. The old allophone string is restored back to the editing line.

ALLOPHONE DATA

The following keys are used:

Cursor keys - Moves allophone cursor
"D" - Displays data on data line for allophone under allophone cursor

The information required to select a particular allophone is contained in Table 2, and may be displayed on the data line. For example, you may wish to know which DD allophone to use in the word "data". Move the allophone cursor over DD1.

Press the "D" key. The following information should be displayed on the data line.

70ms final position: playeD, enD

Now move the allophone cursor over DD2.

Press the "D" key. The following is displayed:

```
   160ms          initial position: Down; clusters: Drain
```

The word "data" has the DD allophone in the initial position so the DD2 allophone should be selected.

ALLOPHONE SEARCH

The following keys are used:

```
Cursor keys     - Moves allophone cursor
"A"             - Search for the allophone under the allophone cursor
"SPACE"         - Display next allophone string
"RETURN"        - Return to main screen
```

Move the allophone cursor over an allophone of your choice.

Press the "A" key. The main screen will disappear and the first allophone string containing the allophone under the allophone cursor will be displayed. The SPACE bar and RETURN key display the next allophone string and return you to the main screen respectively.

CATALOGUE FUNCTIONS

The following keys are used:

```
"C"             - View the catalogue
"SPACE"         - Display next allophone string
"RETURN"        - Return to main screen
"M"             - Load allophone string on editing line to catalogue
"L"             - Link displayed allophone to allophone string on editing line
"R"             - Delete displayed allophone string from catalogue
"T"             - Save catalogue to tape
```

The catalogue allows storage and viewing of allophone strings. The catalogue is stored as a file on tape so you can store your own catalogue of useful words.

The allophone strings are stored alphabetically under their labels. For example the allophone string "KK3,UH,KK2" (allophone representation of the word "cook") given the label "COOK" would be stored in the catalogue under the C's.

The following allophone strings are stored under the L's:

LAUGH, LEARN, LISTEN, LOADING

These can be viewed as follows:

Press the "C" key.

In response to the prompt "ENTER FIRST LETTER", press the "L" key. The first allophone "LAUGH" is displayed.

Press the SPACE bar. The next allophone "LEARN" is displayed.

Press the SPACE bar until all the strings have been viewed and you are returned to the main screen.

Enter the allophone string "LL,AY,PA3,KK2" onto the editing line (allophone representation of the word "like").

Press the "M" key

In response to the prompt "ENTER STRING LABEL", enter "LIKE" and RETURN.

The allophone string "LIKE" is now entered into the catalogue. If you view the catalogue under the L's you should find the "LIKE" entry is placed between the "LEARN" and "LISTEN" entries.

Enter the catalogue again under the L's so "LAUGH" is displayed.

Press the "L" key. This functions links the displayed allophone to the editing line which should now read "LL,AY,PA3,KK2,LL,AR,FF". This function is useful for compiling phrases, but remember that words require a pause between them.

Enter the catalogue again under the L's so "LAUGH" is displayed.

Press the "R" key.

If you now view the catalogue under the L's you should note that the "LAUGH" string has been deleted from the catalogue.

Finally the "T" key will allow you to store the catalogue to tape.

The file provided with the VOXBOX is not intended as a catalogue of useful words but as a catalogue of allophone strings which illustrate different allophone constructions. As soon as you become familiar with these then you can load your own catalogue at the beginning of the program.

BREAK KEY

The break key will cause the program to reset itself. This key should be used if you want to load another catalogue but it should be remembered that the existing catalogue contents will be lost, so should be saved to tape if you want to use them again.

WRITING SOME WORDS OF YOUR OWN

The following is a list of words to try and speak:

CHIN, HOME, ENTER, BEGIN, DARK, SUNSHINE, ERASE, FEATURE, SPEAK, LOADING, FISHERMEN, NOISE, ZOOM, SMILE, ANGER, COMPLETE, SISTER, SPACE, DOORMAT, ELECTRON, LABEL, MISTAKE, INJURE.

First break the words down into their constituent sounds and then use the "allophone data" to help you select which allophone to use where. Enter and speak the allophone strings and when you think that they sound right, compare them with our suggested allophone representations of the words stored in the catalogue.

USING SPEECH IN YOUR OWN PROGRAMS

Side B of the software cassette contains a 2K block of machine code which extends your BASIC with four new commands. It also contains an allophone assembler which allows you to write allophone strings in symbolic rather than numeric format. This code is stored from &D00 to &1500 so no data should be written into this area. PAGE is therefore raised to &1500.

The program is loaded with the command CHAIN"SPEAK" and when it has finished loading the normal power up screen will be displayed with the statement "Extended BASIC" replacing "BASIC". The new commands are SPEAK, LIB, LAST and SSAVE.

SPEAK - This instruction is used to output allophone strings to the VOXBOX. It is similar in format to the PRINT instruction.
Example: SPEAK "LL,AY,KK2";A$,B$

The semi-colon joins two allophone strings together with no pause between them. The comma gives a 400ms pause between strings.

LIB - This instruction is used to display the library of allophone symbols.
Examples: LIB

LAST - This instruction is used to repeat the last allophone string outputted to the VOXBOX.
Example: LAST

SSAVE - This instruction is used to save the extended BASIC code along with your own program.
Example: SSAVE"MY_PROGRAM" or SSAVE A$

A loading header program is saved to tape first. This should be run with the command CHAIN.

If any mistake is made when using the SPEAK command, i.e. entering an allophone symbol incorrectly, the VOXBOX will respond with the error word "Mistake".

Note: The VOXBOX generates NMIs (Non Masking Interrupts) during speech. These interrupts cause the central processor to steal the RAM from the video generator in the ULA. This will cause snow on the screen in Modes 0-3 while the VOXBOX is speaking.

The VOXBOX is designed and assembled in the United Kingdom