

SPEECH CONSTRUCTION SET

OWNER'S MANUAL

FOR USE ON THE APPLE II+, IIe, IIc
with the
COVOX VOICE MASTER
(II+ requires 64K and paddle adapter)

This manual is meant to be used in
conjunction with the
Voice Master User Manual
dated November 1986 or later.

Copyright 1986 by Covox, Inc.
All Rights Reserved.



COVOX INC.

675-D Conger Street • Eugene, Oregon 97402 • U.S.A.
Area Code (503) 342-1271 • Telex 706017 (Av Alarm UD)



TABLE OF CONTENTS

INTRODUCTION.....	1
Software Licensing.....	1
Equipment Required.....	1
Notes On Good Recordings.....	2
Acknowledgements.....	2
PROGRAM CONFIGURATION.....	2
GETTING STARTED.....	3
FIRST LEVEL SPEECH EDITING.....	3
Help Screens and Entering Speech.....	3
Display Modes and Cursor Movement	4
Speech Playback.....	5
FIRST LEVEL MAIN MENU.....	5
Saving and Loading Speech, DOS, Going to Level Two.....	6
Changing Drives and Quitting.....	7
TRANSFERRING SPEECH.....	7
Direct Transfer.....	7
Macros.....	8
Saving Data.....	8
SECOND LEVEL EDITOR.....	9
TECHNICAL LIMITATIONS.....	10
Input Buffer Limitations.....	10
Input Filter.....	11
Second Harmonics.....	11
Voiced vs. Unvoiced Speech.....	12
View Option.....	12
EDITING TECHNIQUES: PART ONE.....	13
Time Expansion.....	13
Time Compression.....	14
Pitch Inflection.....	14
Backwards Speech.....	15
EDITING TECHNIQUES: PART TWO	
SPEECH ENHANCEMENT: THE PASTE AND REPEAT METHOD.....	15
Example One: Improving Vowel Quality.....	16
Example Two: Improving "Saw".....	16
ISCELLANEOUS EDITING TECHNIQUES.....	17
LINKING SPEECH FILES.....	18
Single Word Linker.....	19
Multi Word Linker.....	19
Relocator.....	20

APPENDICES

A. VOICE MASTER SPEECH WAVEFORMS.....	21
B. FIRST LEVEL EDITOR COMMAND SUMMARY.....	21
C. SECOND LEVEL EDITOR COMMAND SUMMARY.....	23
D. FIGURES.....	25

INTRODUCTION

The Speech Construction Set is a new and powerful tool for creating and editing digitized speech. It was originally developed for use in research and has since been enhanced with powerful editing features that are easy to use and understand. The Speech Construction Set is probably the most advanced speech development workstation ever released for any personal computer. Multiple screens and high resolution graphics allow you to "cut and paste" pieces of your digitally recorded speech into a buffer for the creation of high quality speech vocabularies for use in user written talking software. The Speech Construction Set also serves as a valuable teaching tool for learning about human speech--mankind's basic form of communication.

Like any other sophisticated computer aided design (CAD) tool, it takes practice to learn how to use the Speech Construction Set effectively. It also helps to understand some basic concepts of human speech, some of which are covered briefly in Appendix A.

The Speech Construction Set program is not copy protected. BE SURE TO MAKE A BACKUP OF THE MASTER DISK. Replacement disks are available for \$10. In order to receive a replacement disk, or software updates, you must return your original disk.

Appendix D contains all of the numbered illustrations that are referenced throughout this manual. In addition, the Voice Master User's Manual Version 4.0 dated November 1986 or later is referenced frequently and should be in your possession.

SOFTWARE LICENSING

Software and hardware created by Covox, Inc., are protected under United States and international copyright and patent laws. You may use the Speech Construction Set to develop speech for use in your own programs that are not for sale or promotion, but you must clearly state in the manual and in the title screen of your program the following:

Speech developed with the "Speech Construction Set"
available from Covox Inc., 675 Conger St., Eugene, OR 97402

If you write programs for sale or for promotional purposes that use speech developed with the Speech Construction Set, contact Covox, Inc., for licensing information.

EQUIPMENT REQUIRED

In order to use the Speech Construction Set, you must have a Voice Master speech digitizer and microphone. Earlier models, i.e. those encased in a metal box, are incompatible with the Speech Construction Set. Contact COVOX concerning a trade-in policy.

The Speech Construction Set functions with the Apple IIe, IIc, or II+ with at least 64K RAM. The Apple IIGS must be set for IIe mode with a clock speed of 1 Megahertz. An optional sound output card, the Sound Master, is available from your dealer or direct from COVOX. With a Sound Master card installed, sound is produced by either the internal speaker, or by an external speaker or headphone

connected to the speaker jack. The Speech Construction Set allows you to play back speech with or without the Sound Master. We recommend the Sound Master for highest quality speech. You cannot use a Sound Master on an Apple IIc. For the Apple IIc, it is recommended that you attach an external speaker to the earphone jack (or use the earphone on the Covox Headset).

NOTES ON MAKING GOOD RECORDINGS

Professional recording studios and radio stations are keenly aware of the importance of a good voice for intelligibility. Just as important is how well the speaker is able to clearly pronounce each and every syllable (without sounding mechanical). If you are writing commercial software for sale and intend to use voice output, then pay special attention to the selection of the speaker's voice; it should be clear, concise, and robust. If speaking English, the "American" accent used by the national networks or talk show programs is best. Avoid regional dialects unless that is your intent. You may wish to try several voices and pick the one that sounds best. Also, make sure your recordings are done in a reasonably quiet environment.

ACKNOWLEDGEMENTS

The Speech Construction Set and the Voice Master were designed by Brad Stewart. Software for the Speech Construction Set was written by Donald X. Vaccarino. We are interested in hearing your comments or any information you may wish to exchange. Please do so in writing as engineers and programmers tend to work odd hours and can't always answer the phone.

PROGRAM CONFIGURATION

The Speech Construction Set consists of a main program segment and several utility programs which form a complete development system. There are two levels of speech editing. The First Level Editor is the most complicated and is where you begin. You have two speech buffers to work with. The first holds the "raw" recorded speech. The second buffer is where you "construct" speech by selectively transferring portions of speech from the first buffer to the second buffer using "cut-and-paste" methods.

The Second Level Editor converts the speech from the First Level into a compact format that can be spoken by the Voice Master software. It also lets you "fine tune" the amplitude values of your speech. The Second Level Editor is very similar to the Word Editor program that comes as part of the Voice Master software package. It will be quite helpful if you are already familiar with the Word Editor.

The First and Second Level Editors create one-word Voice Master vocabulary files that can be played back using the standard Voice Master playback routines. However, one-word vocabularies are not very practical, so included with the Speech Construction Set are programs that let you link together these one-word files into larger vocabularies.

Other programs on this disk include the latest release of the Voice Master speech recording and playback programs, as well as "stand-alone" playback routines, including a version that will run under Apple ProDOS (see Appendix 5

in the Voice Master Owner's Manual).

GETTING STARTED

Before proceeding, make sure that the calibration and level settings on your Voice Master are adjusted properly. Refer to the section entitled "Calibration and Microphone Technique" in the Voice Master User's Manual. Also make sure that the CAPS LOCK key is pressed down (except for II+).

Plug your Voice Master into the joystick port of your computer. The Speech Construction Set will not run unless a Voice Master is connected to the joystick/mouse port. If you own a II+, you will need a joystick adapter which is available from Covox. Place the Speech Construction Set disk in drive 1 and turn on the computer. If DOS is already present, type BRUN SEBOOT. After a few moments you will be asked if you have a Sound Master installed. (No question will be asked if you own a IIc.) If you do have a Sound Master installed, you will next be asked for the slot number. (NOTE: If you are using a Sound Master with a LASER 128, always specify slot number seven.)

The program next asks for the type of speech it is going to process: male or female. This sets up the input voice pitch filter to an appropriate setting so that second harmonics in your voice pitch are reduced. (This topic is covered in more detail in a later section.) If your voice pitch is high, or you intend to record using a higher pitched voice, select the female voice option.

When the title page appears, you are ready to begin. Press any key and the First Level Main Menu appears. We will begin by going directly to the First Level Editor and demonstrate some of the basic editing features. Then we will return back to the Main Menu and discuss the other options available.

FIRST LEVEL SPEECH EDITING

The First Level Editor is invoked by selecting Option 1 from the Main Menu. To make a selection, press the number desired, or use the up/down or left/right arrows so as to "high-light" the selection, then press RETURN. The First Level Editor screen should resemble Figure 1.

HELP SCREENS

While in the First Level Editor, a list of edit commands can be viewed by pressing the ESC key. Four screens list all of the available commands. Press any key to go from one screen to the next until you return back to the Editor. These Help Screens serve as a quick reference guide to assist you in remembering all the edit commands. Each command will be explained in the examples in this manual. A complete list is presented in Appendix B.

ENTERING SPEECH

It is best to explain the complex editing features of the Speech Construction Set with examples. We shall begin with a simple example to let you become acquainted with some of the basic editing commands. Make sure your Voice Master is connected properly and you have the correct gain setting. In a quiet room, place the microphone within one inch of your mouth. Press the "R" key to start recording and say the word "saw" and nothing else. Recording begins as soon as you press the "R" key and continues until the buffer is full. This takes approximately four seconds and during that time, the text window at the

bottom of the screen flashes. When the recording stops, the program builds up a table of pointers and after a few moments, displays the voice pitch patterns on the top half of the screen.

DISPLAY MODES

Your screen should resemble Figure 2. The dots correspond to pitch frequency. The higher the dot is on the screen, the higher your voice pitch. The dots should appear "in-a-row" during "aw" in "saw" which signifies that it is voiced. The "ss", or "unvoiced" part of "saw" will appear jumbled or random and not "in-a-row". In some cases, the unvoiced portion may be represented with no dots visible but can be identified using the amplitude display (see next paragraph) or using the "View Option" which is explained later on. (The importance of identifying voiced and unvoiced parts of speech is covered in more detail in later section.)

Besides voice pitch, the Speech Construction Set can display the volume level, or amplitude, of your speech. To see a plot of speech amplitude, press the "D" (Display mode) key. If you press "D" again, then both pitch and amplitude are displayed. Press "D" again and only pitch is displayed. Each time you press the "D" key, the display alternates between pitch, amplitude, and pitch/amplitude. The current display mode status is indicated in the second line of the text window.

CURSOR MOVEMENT

There are two cursors for each of the top and bottom buffers. Unless there is some speech placed into the bottom buffer, the bottom cursors will not be visible and cannot be moved. (How to transfer speech to the bottom buffer will be covered shortly.) The cursors are identified by vertical lines at either side of the screen. Only one set of cursors can be moved at a time. To move the cursors in the top screen, the top buffer must be active. The active buffer status is indicated in the second line of the text window. It should read "BUFFER=TOP". To switch buffers, press the "B" key to alternate between the top and bottom buffers.

The right cursor is very important as all editing is done by reference to this cursor. From now on, the right cursor will be referred to as the Edit Cursor. The Edit Cursor is moved by the four keys: H, J, K, & L. To move the Edit Cursor to the left, press "J". High speed cursor movement to the left is accomplished by pressing the "H" key. In a similar fashion, use the "K" for normal, and the "L" key for rapid cursor movement to the right. The left cursor is moved exactly the same way as the right cursor except you must hold down the CONTROL key. If the Edit and left cursors "collide", they will disappear. This can be remedied by simply moving the Edit Cursor to the right or the left cursor to the left.

Each cursor points to a particular amplitude value and pitch value. The values for each cursor are displayed at the bottom line of the text window. There are two sets of numbers which correspond to the values at each cursor. As you move the cursors, these values will change accordingly. The values displayed are not meant to correspond to established units such as decibels or frequency. They are units used by the Speech Construction Set and are useful in determining relative pitch and loudness. The numbers corresponding to amplitude range from 0 to 15 and indicate the 15 amplitude values that are sampled by the Voice Master. The pitch values range from 1 to 189 but only the

values in the range of 17 to 93 are displayed. Each pitch unit corresponds to approximately 130 micro seconds, which is the sampling period. The approximate pitch value in Hertz can be computed by multiply the total number of pitch values by .00013 and then inverting the result. For example, a pitch value of 50 corresponds to $1/(50*.00013)$ or about 154 Hertz.

The top line of the text window signifies the total number of bits, of information that is contained between the cursors. Each bit corresponds to one speech sample. For example, the label, "BOTTOM ACCUM", refers to the accumulated number of samples (bits) between the cursors on the bottom buffer. Divide this number by eight to arrive at an approximate size of the speech data in bytes. Knowing the size of your word is very useful when you must carefully manage your memory requirements for a particular application program.

SPEECH PLAYBACK

There are two playback modes. If you press the "P" key, the entire buffer contents are played backed. If you press the "O" key, only the speech between the cursors is played back.

With a Sound Master installed, you can play speech with it on or off. With it off, speech is produced using the Apple speaker circuit. An on/off indicator is displayed on the third line in the text window which informs you of the status of the Sound Master playback mode. To toggle this flag on or off, press the "S" key. (Without a Sound Master installed, the flag will always indicate the "OFF" condition and cannot be changed.)

Cursor movement combined with playback allows you to accurately define the start and end of a word (or endpoints) in the buffer. When identifying the speech endpoints by ear, it is recommended that you turn the Sound Master off (if equipped) so that the weaker parts of speech come through at maximum loudness. The pitch markers and amplitude lines both help to visually determine the speech endpoints.

At this point, you should be familiar with how to record a word, move the top cursors, and play back the top buffer. Next we will return to the Main Menu and discuss the other options. To do this, press the "Q" key (for quit).

FIRST LEVEL MAIN MENU

The First Level Edit Menu lists nine choices:

- 1) EDIT SPEECH
- 2) LOAD SPEECH INTO TOP BUFFER
- 3) LOAD SPEECH INTO BOTTOM BUFFER
- 4) SAVE SPEECH FROM TOP BUFFER
- 5) SAVE SPEECH FROM BOTTOM BUFFER
- 6) GENERAL DOS COMMANDS
- 7) CONVERT BOTTOM BUFFER
- 8) EXIT
- 9) CHANGE DRIVE

To make a selection, press the number desired for the proper choice, or use the up and down, or left and right arrows to high-light your choice, and then press RETURN.

EDIT SPEECH

Option 1, EDIT SPEECH, puts you into the First Level Editor as mentioned in the previous section.

SAVING AND LOADING SPEECH DATA

Options 2, 3, 4, & 5 allow you to save and load your speech buffers to disk. There are two speech buffers available in the Speech Construction Set. They are labeled top and bottom which refer to top and bottom halves of the screen display while in the First Level Editor. The top buffer is always the buffer that contains the un-edited speech patterns that are recorded by the Voice Master. The bottom buffer is your edit work space. Data from the top buffer is pasted to the bottom buffer by the editing commands in the First Level Editor.

When saving or loading speech data, the contents of each buffer are saved and loaded as three separate disk files corresponding to the speech amplitude, frequency, and pitch. For example, if you save a speech file with the name VOICE, then the disk will contain three speech files named VOICE.A, VOICE.P, and VOICE.C. Do not add the filename extensions when specifying a filename; this is done automatically.

The three speech data files from the top buffer require 16 disk sectors total for storage. The top buffer saves the "original" recorded speech including silence. When you create an edited speech file in the bottom buffer, it is usually much shorter in length than the top buffer. This is because only the information that you actually transfer from the top buffer to the bottom is saved to disk. It is not absolutely necessary to save buffer data, although it is probably a good idea because whenever you exit the First Level Editor and proceed to the Second Level, all your original buffer data, both top and bottom, are erased from main memory.

If a DOS error occurs while accessing the disk, an appropriate message will be displayed at the bottom of the screen and you will be asked to press a key.

ACCESSING DOS

Option 6 lets you use any valid DOS 3.3 command. This is especially useful for formatting blank disks using the INIT command. You can use a drive number specification after the command (e.g. ",D2"). If you type in an invalid command, nothing will happen. To return to the menu, press RETURN right after the prompt followed by any key. Refer to a DOS 3.3 reference manual for more information on DOS commands.

GOING TO THE SECOND LEVEL EDITOR

Option 7, CONVERT BOTTOM BUFFER, will convert the edited speech file in the bottom buffer into the Voice Master format and place you in the Second Level Editor. This will be discussed in more detail in a later section. As mentioned earlier, all data in both the top and bottom buffers will be permanently lost unless they are first saved to disk. You will be warned about this before it's too late at which time you can elect to return to the main menu.

ENDING IT ALL

Option 8, EXIT, will leave the Speech Construction Set and place you back into BASIC. No speech programs are active.

CHANGING DRIVES

Option 9, CHANGE DRIVE, toggles the active drive number between one and two. The active drive is shown at the bottom of the screen. You can also change the active drive by using a drive specifier in a DOS command (refer to Option 6). In a two drive system, you will generally have the Speech Construction Set system disk in drive one and your data disk in drive two.

TRANSFERRING (PASTING) SPEECH FROM THE TOP BUFFER TO THE BOTTOM BUFFER

At this point, you should be familiar on how to record speech, move the cursors, play back speech, and save and load buffers. Now we will explain how to transfer sections of speech from the top buffer to the bottom buffer. We'll use the "saw" example from a previous section. Speech quality enhancement will be discussed in a later section.

DIRECT TRANSFER

In this first example, we will simply transfer a portion of the top buffer, in this case, the word "saw," to the bottom buffer. No special editing or speech enhancement methods will be used. Let us assume that you've already recorded the word "saw" from the previous example and have determined the endpoints using the cursors. Move the Edit Cursor (the right one) to the beginning of the word (next to the left cursor). Now if you press the "I" key, all the data contained within the pitch period pointed to by the Edit Cursor (including the amplitude) is transferred, or pasted, to where the bottom Edit Cursor is, in this case, at the beginning of the bottom buffer. One dot should appear at the bottom buffer. Next, press the "K" key to move the top Edit Cursor one dot to the right and press the "I" key again. At this time, the bottom cursors should appear. By following the key sequence, IKIKIK, etc., the bottom buffer starts taking on the appearance of the top buffer. You can stop when the top Edit Cursor has reached the end of the word.

By now, you should have successfully transferred the entire word "saw" to the bottom buffer (Figure 3). To hear how it sounds, you must press the "B" key which "activates" or transfers all cursor and playback movement to the bottom buffer. The current active buffer is indicated in the text window and should read, "BUFFER=BOTTOM." If it doesn't, press the B key again. At this point, press "O" or "P" and you will hear "saw" spoken. It should sound identical to "saw" in the top buffer.

If the top window contains several words, or perhaps a sentence in which your chosen word "saw" occurs, then you can transfer only this one part to the bottom buffer. Do this by first moving the left cursor to where you want the transfer to begin and then begin the Edit Cursor next to the left one as has already been described.

MACROS

As you may have noticed, pressing 1K1K1K, etc., is somewhat time consuming. The Speech Construction Set allows you to define and use one macro command at a time (but you can change macros while editing). This can save you a great deal of time and effort. Let us repeat the above example, but this time we'll use a macro. First, erase the bottom buffer by pressing CONTROL-E (Erase) and answer the prompt. (CONTROL-E does not erase the top buffer.) Next, press "B" to select the top buffer (if necessary). Move the Edit Cursor to the beginning of the word as before. Then define the macro as follows: Press "Y" and enter "1K1K1K1K" and press the return key. The macro you have just defined replaces eight key strokes and is displayed in the bottom text window. To use the macro, press the "U" key (for Use). Each time you do this, four pitch periods will be pasted and the Edit cursor will move to the right by four periods. However, you will not see the graph being built up on the bottom screen until you enable the bottom buffer with the "B" key, or if you paste without using the macro. (Updating the bottom display when using macros would slow down the data transfer by a noticeable amount.)

SAVING DATA

Once you are satisfied with the contents of the bottom buffer, you can return to the main menu by pressing "Q". We will continue this example all the way through to the Second Level Editor. Later, we'll return to the First Level Editor and work with the same word, "saw", but we'll edit it differently each time.

But before we enter the Second Level Editor, let us save the contents of the top and bottom buffers. Place a blank, formatted disk into drive 1 (or into drive 2).

Note: To format a disk, place a blank diskette into drive 2 (or drive 1 if you have only one drive). Then select GENERAL DOS COMMANDS from the Main Menu. Then enter the command, INIT HELLO,D2 (or INIT HELLO,D1 if using drive 1). After about a minute, your diskette is formatted and ready to accept data. Press RETURN twice to return to the Main Menu.

Select Option 4, SAVE SPEECH FROM TOP BUFFER. Use a filename of your choice, such as "SAW-TOP". Next, save the bottom buffer contents with Option 5, SAVE SPEECH FROM BOTTOM BUFFER using a filename such as, "SAW-BOTTOM."

Now select Option 7, CONVERT BOTTOM BUFFER. The program will first warn you that you are about to erase all your buffer data. You can elect to return to the menu and save your buffers (if your forgot to). Otherwise, the program prompts you to put the Speech Construction Set disk into drive 1 and press RETURN.

Note: It is not mandatory to save the contents of the bottom buffer before converting it. But if you don't, you won't be able to re-edit the bottom buffer at a later time if the results of the Second Level Editor are unsatisfactory. If you spent a lot of time editing, then it is probably a good idea to save the efforts of your labor. As you gain experience with the Speech Construction Set, you

may find that it no longer becomes necessary to save the buffers each time to enter the Second Level.

SECOND LEVEL EDITOR

When you elect to go to the Second Level Editor (by selecting Option 7, CONVERT BOTTOM BUFFER), the entire contents of the bottom buffer are converted into the Voice Master format used in speech playback. The ability to manipulate individual voice pitch periods is irretrievably lost (unless you saved the bottom buffer to disk). After the conversion is complete, the Second Level Editor Menu appears.

The Second Level Editor, or Amplitude Editor, allows you to adjust the amplitude or loudness levels of the speech every 20 milliseconds. Most of the commands are identical to the Amplitude Editor described in more detail in the Voice Master Owner's Manual. The main differences are: 1) that you can only edit one-word vocabularies and, 2) you can toggle speech playback between the Sound Master and the Apple speaker. A list of Second Level editing commands is given in Appendix C.

Speech is recorded with 15 levels of loudness. There are 15 levels of loudness available for playing back speech, but only if the proper playback hardware is installed, such as the optional Sound Master card. Without a Sound Master card, you can reproduce speech at only two levels--on and off. Therefore, adjusting the amplitude has no audible effect without a Sound Master. However, we have developed some software "tricks" to make your ear perceive differences in loudness even though only two levels are available.

NOTE: The 15 levels reproduced by the Sound Master are not uniform. This is a characteristic of the chip used in the Sound Master (GI AY-3-8913). Amplitudes below 7 drop off very quickly. This limits the apparent range of loudness. Therefore, it is advisable to keep most of the amplitude values between 8 and 15, except where you want complete silence.

We can view what the converted buffer looks like by pressing Option 1, EDIT A WORD, from the Second Level Menu. What appears is a sequence of asterisks that corresponds to an amplitude value sampled every 20 thousandths of a second (20 milliseconds). The screen will hold 40 of these amplitude samples, or about 8/10 of a seconds worth. The word "saw" should fit within the screen. There is one Edit Bar and it is shown in reverse video. It should be located at the extreme left side of the screen. This bar can be moved left and right by pressing the "J" and "K" keys respectively. As with the First Level Editor, a Help Screen with a list of edit commands can be viewed by pressing the ESC key. Press any key to return to the Editor.

To hear the contents of the buffer, press the "P" key. To hear the section of speech from the left side of the screen to the Edit Bar, press the "O" key making sure that the Edit Bar is positioned towards the center of the screen. If you have a Sound Master installed, you can elect to turn the Sound Master on or off by pressing CONTROL-S. The "s" in "saw" is probably too loud. Let us locate the end of "s" in the word by moving the Edit Bar back and forth and pressing "O" until all you hear is "s" and not the beginning of "aw". Then use the "S" key to quiet down the "s" sound. (Refer to the Amplitude Editor in the Voice Master Owner's Manual for more detailed information on this type of

editing.)

When you are satisfied with how the word sounds, return to the Second Level Menu by pressing the "Q" key from the Editor. At this time you will probably want to save this one-word file. From the menu, select the proper disk drive you want to save your file to (Option 5) and proceed to save it (Option 3). Call it "SAW". Later on, you can link up to 64 of these one-word vocabularies into one large Voice Master vocabulary file using the Linker Utilities discussed in a later section.

NOTE: You may have noticed that the speech quality is better in the Second Level Editor. This is because the playback routine is much simpler and the critical software loops can be time adjusted. The playback routine used in the First Level Editor is extremely complicated and pushes the 6502 microprocessor to just beyond its maximum capabilities. Thus, some of the critical timing loops cannot be adjusted for even sampling.

Now that you have successfully finished saving your first word, return to the First Level Editor with Option 7, RETURN TO SPEECH CONSTRUCTION SET.

SOME TECHNICAL LIMITATIONS

Before going further, let us discuss some of the technical limitations that you should be aware of.

INPUT BUFFER LIMITATIONS

It is important to understand differences between male and female voice pitch and how the Speech Construction Set processes and displays these differences. The female voice pitch is approximately twice as high as a male voice. The male voice will average at around 110 cycles per second and the female at around 200 cycles per second. A young male (under 12 years) will have a voice pitch near that of a female. Small children's voice pitches are even higher.

In the Speech Construction Set First Level Editor, voice pitch is displayed as a series of dots on the screen. The vertical axis represents the pitch value; the higher the pitch, the higher the dot is on the screen. The horizontal axis displays the pitch sequence in time from left to right. However, the horizontal axis does not directly indicate how long the word is. It takes more horizontal space to display a higher pitch value for a given word length. For example, let us suppose that a pitch period exists that is 50 time samples long as indicated on the bottom line of the text window. If we had 10 of these time samples in a row, the horizontal axis displays 10 pitch dots. The total length of the word, in time samples, is 10 times 50, or 500 time samples. If the pitch is now raised by an octave, i.e. twice as high, then that pitch period is only 25 time samples in length instead of 50. Ten of these shorter pitch periods total 250 time samples. We can see that 10 periods of a 25 unit pitch value is half as long in time as 10 periods of a 50 unit pitch period, even though the length of the horizontal display is the same. Figure 5 shows two repetitions of the word "one," first with a male voice and then with a female voice. Notice that the female voice is about twice as high in pitch and the display is about twice as long, even though it took approximately the same

amount of time to speak the word in both cases.

INPUT FILTER

The input filter will combine or add together two pitch periods into one longer value before displaying them to the screen. The main reason why you may want to use an input filter is to reduce second harmonics in the voice pitch (discussed in next section).

The pitch periods that are selected for filtering are those that exceed a user defined filter value. This filter value can be changed by pressing the "F" key for Filter. The filter values range from 0 to 80, where 0 turns the filter off, and 80 maximizes the filter effect. The default values are 45 if you selected the male voice option or 0 if you selected the female voice option when the program is first run. A filter value of 45 means that any pitch period with a value less than 45 will be combined with the pitch period immediately following it. If the filter value is set too high, then higher pitch values will "fold over" as shown in Figure 4. "Fold over" on voiced sounds will hamper methods for reducing distortion (to be discussed later on).

The higher voice pitch of a female or child can pose a problem with the Speech Construction Set. As was discussed previously, it takes more horizontal space to display a higher pitch value for a given word length. As a result, you may find out that the word or phrase you want to edit extends beyond the edge of the screen. Since all you have available is one screen of input space, you have only one option available: speak faster and use shorter or fewer words.

SECOND HARMONICS AND FILTERING

Sometimes the pitch value that you select for pasting may not be a good choice. The pitch tracking circuit inside the Voice Master can cause errors. These show up as a sudden rise of the pitch caused by a second harmonic being generated in the pitch tracker. The software thinks your pitch is twice as high as it actually is. Figure 6 illustrates this problem in more detail. Notice that the second harmonic pitches occur in pairs and are approximately half the value they should be, i.e. a second harmonic of a pitch value of 50 results in two periods that total 50--one may be at 23 and the other at 27. Selecting a second harmonic period for transferring to the bottom buffer will cause errors.

Second harmonics are generally present in male voices and especially with certain voiced sounds, such as "oo" as in "too" and "ee" as in "speech". With some male voices, the second harmonic situation arises so seldomly that you can "skip over" them with the edit cursor and pick a "good" pitch period next to it. However, there are many cases where all you have is second harmonic pitch. Thus you need a method of "filtering out" the second harmonic. The best way to do this is to set the "pre-filter" value as discussed in the previous section. However, it is possible to remove the second harmonics after the speech has been recorded. The male filter default value of 45 will generally remove most second harmonics. If not, try increasing the value slightly.

The Speech Construction Set has two filtering methods to remove the second harmonics after the speech has already been recorded. You may need to use these filtering routines if the pre-filter setting was too low to filter out second harmonics. The first routine filters a single period (CONTROL-F) and another filters all periods between the cursors (CONTROL-R). As with the pre-filter,

these two post-filtering algorithms operate only on the speech in the top buffer. Figure 7 shows the filtered results of Figure 6 using CONTROL-F.

A second harmonic can be identified as two dots higher than the "normal" level of pitch dots (Figure 6). The first dot of the pair is at least equal to and generally higher in pitch than the second dot. When using the single-dot filter, CONTROL-F, always place the right cursor at the first dot; then press CONTROL-F. The two dots will "merge" into one lower pitched dot whose value is consistent with what would be "normal". When filtering a range of second harmonic dots (CONTROL-R), place the cursors so that they surround the area to be filtered. The post-filter value (CNTL-R) is the same value as the input filter setting and can be changed by pressing "F".

VOICED VS. UNVOICED SPEECH

By definition, a voiced sound is that which is created by the larynx, or "voice-box". These include all vowel sounds ("ah, ee, oo"), nasals ("m", "n"), certain plosives ("b", "g") and voiced fricatives ("v", "z"). In short, any time your voice-box is vibrating, you are producing a voiced sound. What happens is that each time your voice-box vibrates, it "shock excites" the oral cavity of your mouth. In other words, your mouth "rings" each time it is "struck" by your larynx. How your mouth is shaped determines the sound quality. Try saying "ee" "oo" and "ah". Notice how different the shape of your mouth is for each of the three vowels. Close your mouth, and you produce a nasal sound like "m" or "n".

By contrast, unvoiced sounds do not use the voice-box for reproduction. These sounds include the sibilents ("ss", "sh"), plosives ("t", "k", "p", "ch"), and unvoiced fricatives ("f", "th"). In English, unvoiced sounds are generally produced by air moving across the lips. Whispered speech is also considered unvoiced, except that the source of the sounds is in the throat for sounds that would normally be voiced. Most unvoiced sounds are acoustically similar. They differ primarily in loudness, duration, and envelope shape. For example, if you take an "s" sound and shorten it, it becomes a "t". But they all share a common trait--the sounds produced consist of random "Gaussian" noise.

It is very important to understand the basic difference between voiced and unvoiced sound. Creating high quality speech requires that you identify the voiced and unvoiced parts of speech sample you are editing. The general "rule-of-thumb" for identifying a voiced sound is to note the regions where the pitch dots are strung together "in-a-row". Figure 3 illustrates this point. Some exceptions to this rule are those sounds that are a combination of voiced and unvoiced sounds, such as "v" and "z".

THE VIEW OPTION

Each pitch period contains a group of higher frequency square waves. These square waves represent the "ringing" of the oral cavity when "shock excited" by the larynx at the pitch rate. To see what these square waves look like, place the Edit Cursor on a pitch period and press "V" for view. The text window at the bottom of the screen is replaced by a graphics window depicting the high frequency square waves (Figure 7). The higher the pitch period, the shorter the group of square waves. For long pitch periods, the square-wave extends beyond the length of the screen and "wraps around" to form another line beneath the first. Each group of square waves can be listened to by pressing the "C" key. Pressing "C" takes the selected pitch data and repeats it ten

times in a row. In this manner, you can identify how each of the group of square waves in a pitch period contributes to the overall word.

When you invoke the View Option, you can also alter the square waves themselves. You will note that an edit bar exists just below the start of the square-wave set. This edit bar can be moved left and right by the use of the "Z" and "X" keys, respectively, or moved at high speed with CONTROL-Z and CONTROL-X. The space bar is used to change the state of the square immediately above the edit bar. For example, if the square wave above the edit bar is "low" or in a zero-state, pressing the space bar changes it to a "high" or one-state. Pressing the space bar again returns it to the "low" state.

Viewing the square wave group is useful in many respects. You can visually, as well as aurally, select those pitch periods that sound best. You can select silence periods or convert periods into silence. (A silence period is identified by noting that the square wave group is at one level only, i.e. it remains in either a high or low state with no transitions between levels.)

Having the ability to alter the pitch period waveform presents some interesting possibilities. Small alterations can dramatically affect the characteristics of the voiced sound. Putting in more one-to-zero transitions towards the beginning of the period can make a vowel such as "ah" sound more like "ee." (You can also do the reverse.) By blanking the last third or half of a period, you can add a low frequency component to your speech which can give it a "deeper" voice quality, or "sharpen" the fundamental pitch of the talker.

EDITING TECHNIQUES

PART ONE

The previous example on transferring speech from the top to the bottom buffer was simple; very little editing was really accomplished. The primary purpose of that exercise was to familiarize yourself with the pasting method and with the editing functions. In the following examples, we will examine methods for expanding, compressing, and inverting the time base information of the speech waveform. The following are given primarily to further your understanding of the editing process which is a prerequisite in PART TWO of EDITING TECHNIQUES in which method for reducing distortion are discussed.

TIME EXPANSION

Time expansion lengthens or expands the time it takes to play back speech. For example, if we expand the time by a factor of two, then a word that originally took 1/2 second to play back now will take a full second. The simplest way of accomplishing this is to slow down the playback rate by 1/2. But this also lowers all the speech frequencies by 1/2, which seriously degrades intelligibility. Another approach is to "stretch" the waveform to twice its length. This is the approach we'll use in the following example.

Start by loading the top buffer with the speech patterns of the word "saw" that you recorded earlier and saved. If you did not save the original buffer, simply re-record it. We will now create new bottom buffer data so erase any existing speech in the bottom buffer with CONTROL-E.

Define the start and end portions of the word using the cursors. Next,

place the Edit Cursor next to the left cursor at the beginning of the word. This time, use the key sequence "2K2K2K", etc., to paste each period from the top buffer twice to the bottom buffer and then move the Edit Cursor to the right by one dot. As you continue to paste, the bottom buffer will appear to look like the top except it will be twice as long as in Figure 8. You can define a macro to do this in order to save yourself some time. After you have pasted in the entire word, switch buffers and hear what you have created. It should sound the same but will take twice as long to play back. (You may notice that the "s" in "saw" has a slight mechanical quality to it. This is because you are pasting multiple periods of an unvoiced sound with is something to avoid and will be discussed in the section on speech enhancement.)

Time expansion is a useful means of emphasizing a portion of speech in a word. Let us suppose that we recorded the word "density" and the "n" is not pronounced clearly. You can augment the sound by pasting a few more pitch periods of "n" into the bottom buffer in the appropriate place. (Speech data is always inserted to the right of the Edit Cursor in the bottom buffer.)

TIME COMPRESSION

Time compression is a way of speeding up speech by compressing the time it takes to play back speech. This can be accomplished by speeding up the playback rate, much like playing a 33 RPM record at 45 RPM. But this raises all speech frequencies and the result sounds more like a chipmunk than a human. A better method is to remove portions of speech without leaving gaps. This is easily accomplished by pasting a smaller number of pitch periods from the top to the bottom buffer. For example, to compress speech down to one half of its original length, paste using the key sequence, "1KK1KK1KK", etc. In other words, paste once, move twice. (Again, a macro would be useful here.) If you do this, the resulting speech pattern will appear half as long as the original, as in Figure 9. Before trying this, erase the bottom buffer using a CONTROL-E. In a similar fashion, pasting three out of every four periods reduces the playback time by three fourths.

Time compression is useful for shortening parts of speech. For example, the "s" in "saw" can be shortened in duration in this manner. (An interesting experiment to try is to make the "s" so short that it sounds more like a "t" or "k" sound. This is also explained in the Voice Master Owner's Manual in the section on the Amplitude Editor.) Time compression is useful when you need to create a large vocabulary in a limited amount of available memory--you can record each word or phrase as you would normally and then compress the words. Another way to time compress, which is often more practical, is to delete portions of speech from the bottom buffer after transferring the entire word. This is accomplished by first locating the Edit Cursor to the portion of speech you want to remove and then press CONTROL-D (for Delete). Each time you delete a portion of speech, all of the speech to the right of the Edit Cursor is shifted to the left by the amount deleted. The lower the pitch dot, the more speech you are removing as lower pitches require more memory storage. You can also delete unnecessary parts at either ends of the word.

PITCH INFLECTION

The recording of "saw" that you made probably was spoken in such a manner that the end of the word went down in pitch. But suppose you recorded "saw?" as a question. The end part of the word would raise in pitch which might signify to a listener that a question is being asked. It is possible to manually edit

pitch changes using the Speech Construction Set editor. There are two edit commands to alter pitch: pressing "I" will raise the pitch value at the Edit Cursor by one sample period and "M" does the opposite.

Figure 10 illustrates how the voice pitch values towards the end of "saw" can be adjusted upwards from the original. The main purpose of adjusting pitch is to "smooth out" the approximations to pitch made during the removal of distortion which is elaborated on in the speech enhancement section of this manual.

BACKWARDS SPEECH

The Speech Construction Set editor can be used to paste speech inverted, or backwards. The following example illustrates how it is possible to take a word from the top buffer and paste it in reverse order to the bottom buffer. Begin by placing the top left cursor at the beginning of the word, and the right Edit Cursor at the end of the word. Clear the bottom buffer with CONTROL-E. Then use the key sequence 1J1J1J1J, etc., until the right Edit Cursor meets the left cursor. As you progress, you should begin to notice that a mirror image of the speech pattern in the top buffer starts to take shape in the bottom buffer, as shown in Figure 11. Switch to the bottom buffer and press "O" to hear the result.

Although there may be no practical purpose for pasting speech backwards, except for decoding "demonic rock and roll lyrics", it demonstrates the flexibility of the editor.

EDITING TECHNIQUES: PART TWO

SPEECH ENHANCEMENT: THE PASTE AND REPEAT METHOD

In this section, we will discuss how to improve the intelligibility and clarity of your digitized speech patterns. This is a very powerful method of re-coding speech but it takes a good understanding of the philosophy behind a "paste-and-repeat" approach. At this point, it is assumed that the reader can record speech, manipulate the cursors, play back speech, transfer speech from the top to the bottom buffer, and identify voiced and unvoiced speech.

Sounds digitized by the Voice Master are coded with a one bit analog-to-digital converter. Any portion of speech above a certain voltage threshold is assigned a value of one; below this threshold, the value is assigned a zero. Sampling in this manner produces "infinitely clipped speech". Since a one-bit converter is limited in accuracy, errors are present. When played back through a one bit digital-to-analog converter, the speech, though quite intelligible, has a harsh, raspy quality to it. This distortion is due to several factors, the primary one being differences in the phase characteristics of the clipped speech waveform between each successive voice pitch period. Eliminating these phase differences dramatically improves the speech quality.

The Voice Master hardware and the Speech Construction Set software isolate and identify voice pitch periods so that a simple method of paste-and-repeat effectively filters out much of the phase discrepancy. Let us try a simple experiment to illustrate this. Record about one second worth of "ee" as in "speech." Now play it back to hear how it sounds. There exists substantial distortion. But suppose we examined just one pitch period from "ee" and played

it over and over again. What would it sound like? There exists a special playback mode in the Speech Construction Set which repeats a pitch period 10 times in succession. Move the Edit Cursor to any pitch dot on the display and press the "C" key. It should sound very much like an "ee" but with a slight mechanical quality. To see what the pitch period waveform looks like, press the "V" key to toggle the viewing screen. The text window is replaced by a graphic representation of the waveform. Press "V" again to return to the text window.

Repeating an identical pitch period has effectively removed any phase discrepancies between pitch periods (since it is the sample period repeated several times).

EXAMPLE ONE: IMPROVING VOWEL QUALITY

Prepare the microphone and record, without pausing, the word "away". Erase the bottom buffer. Place the Edit and left cursors at the beginning of the word in the top buffer. Now use the paste sequence, 4KKKK4KKKK, etc. until the entire word is pasted. The result should resemble Figure 12. When you play the result back, it should sound much cleaner. What you are doing is pasting four periods of every fourth pitch period in the top buffer to the bottom buffer. The end result approximates the original in a "piece-wise linear" fashion less the distortion. This, then, is the essence of the paste-and-repeat method of approximating a speech waveform. If you paste-and-repeat only once, you wind up with the original speech with no improvement. If you repeat too many times, you wind up with mechanical (or robotic) sounding speech. A good compromise seems to be between 2 and 6 paste-and-repeats. Generally, the higher the pitch is, the more pitch periods you can paste-and-repeat. There are a few rules that need to be followed. The most important one of these is: You should only paste-and-repeat voiced sound. Unvoiced sounds should be pasted verbatim, i.e. one at a time.

EXAMPLE TWO: IMPROVING "SAW"

In this second example, we will edit the word "saw" using a paste-and-repeat process. The difference here is in knowing that the unvoiced part ("s") must be pasted differently than the voiced part ("aw"). Pasting repeated portions of an unvoiced sound makes it sound non-voiced. The "s" in saw is pasted using the sequence 1K1K1K, etc. This maintains the noise-like qualities of the sibilant "s". (As an experiment, you may want to try pasting "s" like a voiced sound. You'll find out immediately why you can't do this.) The rest of the word should be pasted with the sequence, 4KKKK4KKKK, etc., as this is voiced. You can easily determine where the "s" ends and the "aw" begins by listening to the speech between the cursors by pressing "O" as you move the Edit Cursor to the right. The final result should resemble Figure 13.

If you listen to the constructed version of "saw" as in Figure 13, it should sound pretty clear, especially the voiced portion. (The unvoiced part will be "quieted" in the Second Level. In addition, simplified playback software in the Second Level makes speech sound better.) The main difference in the appearance between the two buffer versions of "saw" is that pitch plot in the bottom buffer is "stepped" as opposed to the smoother curve in the top buffer. This will generally pose no problem if the "steps" are small. However, if the steps are large, which is the case if the pitch glide is rapid, then the sound quality will have a "flutter" to it. This can be alleviated by raising or lowering the pitches at each step in order to "smooth" out the curve using the "I" and "M" keys. This is shown in Figure 14.

PART THREE

MISCELLANEOUS EDITING TECHNIQUES

The following editing techniques are compiled from the experience gained by editing hundreds of words using the Speech Construction Set.

Speech is always transferred from the top Edit Cursor to the bottom Edit Cursor regardless of which buffer is active.

Sometimes it is necessary to insert silence periods between phonemes or words. This is easily accomplished by selecting the View Option (with the "V" key) and locating a silence portion in the top buffer. However, most silence periods are quite long; typically 189 samples. You can create a shorter silence period by selecting a pitch period from a voiced part of the word and then change it to silence by setting the square-wave to all ones (high) using the View Edit Cursor and the space bar.

Use the View Option to locate the beginning of a word that has a weak beginning part. For example, the "th" in the word "three" is very weak and may even indicate a zero amplitude or loudness level. However, by moving the Edit Cursor and looking at the square-waves, the beginning of the word can be identified by when the square-wave changes from silence (one) and begins to alternate between one and zero. Avoid pasting square-waves with a lot of silence in them at the beginning or end of a word as this wastes memory.

While in the First Level Editor, do most of your work with the Sound Master disabled. It is easier to hear the weaker parts of a word this way. With the Sound Master on, you can change the loudness of the word using CONTROL-I and CONTROL-M but it is much easier to do this in the Second Level Editor. (If you decide to define a macro to lower amplitudes, you must use the alternate key for lowering amplitude, which is "N". You cannot use CONTROL-M as the Apple interprets this as a RETURN.)

Certain voice sounds are sometimes difficult to paste properly. Example of these include the voiced fricatives such as "z", "v", or "th" as in "the". Notice that these voiced fricatives are a mixture of voiced and unvoiced sounds, e.g. "z" is a combination of "n" with "ss". Because of the strong unvoiced component, the pitch-tracker circuit in the Voice Master is unreliable which can cause pitch periods to appear random and not "in-a-row" as would be desired. There are two ways to paste these voiced fricatives and which method you choose depends upon the voice used in recording and your best judgement. The first method is to paste each period one-by-one without repeating. The second method does involve some paste-and-repeat, but try to pick random periods that are near the "normal" pitch value for that particular voice and then paste-and-repeat only only two or three times; any more and the result sounds more like a mechanical buzz instead of a voiced fricative. Then, when you go to the Second Level Editor, use the "Z" key to high pass filter the region where the voice fricative is and adjust the amplitude to about 8 or 9.

In the Second Level Editor, you can improve the sound of the "f" fricative sound by using the "X" key several times (two to four) followed by "S" in the fricative region. In a similar manner, the "th" fricative can be improved except use the "X" key a little more liberally (4 to 8 times) over the fricative region finishing with the "S" key.

You can create an interesting effect by pasting a whispered sound like you would a voiced sound. The result is pretty weird.

It is quite possible to record one complex phrase that contains most of the vowels and some unvoiced sounds and use it as a "library" where pieces can be pasted and altered. It should then be possible create just about any word you can think of.

You can make your voice sound like a robot by pasting many voice periods of the same pitch period (more than four) and adjusting the pitch values (using "I" and "M") so that they are all at the same value.

If you need to create a longer unvoiced sound in the bottom buffer than what is available in the top buffer, then randomly select unvoiced pitch periods and paste them into the bottom unvoiced region. If you do this properly, the result will still sound unvoiced.

It is quite possible to create a Voice Master vocabulary that only contains phonemes. Then you could use a suitable "text-to-speech" algorithm to speak any word as it is typed. Although it would sound robotic, the constituent parts of the speech are your own! If you happen to write a program that can do this, we would be interested in hearing from you!

LINKING SPEECH FILES

The Speech Construction Set editors let you create one-word vocabulary files that are in the Voice Master format. They can be read by the &FIND command and played back with &SPEAKO. However, one-word vocabularies are impractical as most applications require vocabularies that are longer than one word. Thus, we have provided a means to link these one-word vocabularies into larger files. In this way, it is possible to have up to 64 different words and phrases in memory at one time. For example, let us suppose you fully edited and saved as a one-word Voice Master file, the word "saw". In addition, you have already created the words "hammer", "chisel", and "screwdriver" and saved all of them as one-word Voice Master files. Now you can combine these one-word vocabularies into a four-word file called "TOOLS".

There are two linker programs provided. The first one allows you to link together one-word vocabularies only. The final linked file always has speech starting at location \$4000, or page \$40 (location 16384 decimal).

The second linker program allows you to link together multiple-word vocabularies linked by the first linker. In addition, the second linker lets you redefine the base address for speech. (Refer to the &RESET command in the Voice Master Owner's Manual.) However, the second linker program can only be used with an Apple II with auxiliary memory, i.e. 128K. If you need to move speech to another location in memory, yet you do not have auxiliary memory, then a third, non-linking, relocater program is available.

To use the linker programs, type RUN LINKER and press the return key. After the program loads and runs, a menu lets you select which of the three programs you want to use.

LINKER1 --Single Word File Linker (64K or 128K).

This program will link together one-word files only. If you try to link vocabulary files with more than one word, an error message will be displayed. You are prompted to enter the filenames of the one-word vocabularies you want linked. Enter them in the desired order. The current word number in the linked file you are building is displayed at the left. If you want to skip a word, then press return without entering a filename. Filenames can include a disk drive specifier, e.g. FILENAME,D2. When you have entered the last single-word vocabulary, press the CONTROL and S keys simultaneously, then press RETURN. You will then be asked for the filename you want this linked vocabulary saved as. Enter any valid DOS 3.3 filename. You can also add the ",D1" or ",D2" drive specifier to the filename. After your linked file has been saved, you are returned to BASIC.

You can exit LINKER1 anytime and return to BASIC by pressing CONTROL-C followed by RETURN instead of entering a filename. This is handy if you forgot one of your filenames and need to catalog your disk. Unfortunately, you will have to restart the program by typing RUN and press RETURN.

LINKER2 -- Multi-Word File Linker (128K Only).

This program allows you to link together multi-word (or single-word) vocabularies that were linked together using LINKER1. You can run this program only if you have a 128K IIe, IIc, or equivalent. LINKER2 first loads the Voice Master programs PARTAEX and PARTBEX, (or PARTAE and PARTBE if you have a Sound Master) so make sure they are on the same disk as LINKER2.

The program then asks you for the page number you want the speech to start at. If your application is to run on an Apple with at least 128K, select page 16 as this allocates most of auxiliary memory for speech. However, if your speech application is designed to work in a 64K environment, then select the page number that is appropriate for your particular application. The higher the number, the less memory is available for speech and you should be aware that the linking of longer speech files may exceed the allocated memory.

There is one minor limitation when using LINKER2 to link together two or more multi-word files. And that is that each vocabulary file must have all its words recorded or linked in ascending order. For example, if you created a file that had words recorded or linked in the order 3,2,1, instead of 1,2,3, LINKER2 will crash. If you link together files that were originally linked using LINKER1, then this problem will never arise.

If you have a 128K Apple, use LINKER2 to relocate a single speech file instead of the RELOCATOR program described below. (The RELOCATOR program was provide solely for the benefit of Apple II+ and IIe owners with 64K memory.) Simply enter the filename of the vocabulary you want to relocate and press RETURN for the second filename.

RELOCATOR -- Relocate a Speech File in Memory (64K).

When you convert speech into Voice Master format from the Second Level Editor, the single-word speech file is always stored beginning at location hex \$4000. Memory location \$4000 is also referred to as "page \$40". This means that all memory from page \$40 up to page \$95 is reserved for speech (see Appendix 2 in the Voice Master Owner's Manual). This allows for almost 30,000 bytes, or 60 1/2 second words. But you don't have much room for a program or graphic screens (under 16K). In other cases, you may have a short speech vocabulary but need lots of memory for your program or graphic screens, so you would like to locate speech on a higher page than \$40.

The RELOCATOR program first asks you for the name of the speech vocabulary file to relocate and then for the page number (14-114) where you want the speech file to load to. If the relocated vocabulary won't fit in the allocated memory, then you will be asked to enter a smaller number. If accepted, the vocabulary is resaved to the same disk drive and under the same filename from which it came, so you should have a backup of this file before proceeding.

APPENDIX A: VOICE MASTER SPEECH WAVEFORMS

The Voice Master hardware converts the sounds picked up by the microphone into three digital signals for computer input. When the input sound is speech, the three waveforms correspond to voice pitch, a high frequency speech carrier, and amplitude.

The voice pitch (also called the fundamental frequency) indicates the rate at which your voicebox vibrates. It typically ranges between 100 and 300 cycles per second. Male voices average around 120 cycles per second and female voices around 200 cycles per second. The pitch frequency is displayed as a series of dots on the First Level Editor screen.

The speech carrier contains most of the "formant" frequencies of speech and convey most of the intelligibility. Hence the name "carrier"--it carries the speech information. The carrier is a high frequency square wave and can be observed using the View Option ("V" key) when in the Level 1 Editor of the Speech Construction Set. Carrier frequencies typically range between 500 and 4000 cycles per second. Another term for the carrier wave is "infinitely clipped speech".

The last waveform contains the loudness or amplitude of the speech. It is a square wave whose duty cycle is proportional to loudness. Amplitude values range between 0 and 15 with 15 being the loudest. The amplitude values can be observed as a connected line plot on the First Level Editor screen and as a series of asterisks in the Second Level Editor Screen.

Figure 15 illustrates how a typical voiced waveform would look like on an oscilloscope. All three of the components of speech are contained within this complex waveform. All three are converted into square waves by the Voice Master. Notice how evenly spaced the pitch waveform is. This is characteristic of a voiced sound. An unvoiced sound, such as "ss", produces random or irregular pitch periods.

APPENDIX B: FIRST LEVEL EDITOR COMMAND SUMMARY

Note: All commands require that the CAPS LOCK key be pressed.

ESC KEY List the Help Screens.

B Buffer toggle. There are two buffers labeled top and bottom. Only one buffer can be active at a time. You select buffers by pressing the B key. Current active buffer is indicated in the text window.

R Record. Pressing this key begins the recording process. The text window flashes while recording. Recording ceases when the input buffer is full which takes approximately four seconds.

Sound Master toggle. This toggles the speech output between the optional Sound Master card and the internal sound generator. Status is displayed in the text window. Not applicable for Apple IIc.

Create macro command. After pressing Y, you will be asked to enter a macro command sequence of up to 10 characters. Any of the editing commands can be entered. The macro is listed in the text window.

Use macro. Performs the editing functions by a user defined macro.

Toggle carrier view. Switches bottom of screen between a text window and a hi-res display of the carrier square waves within a pitch period.

Z Move carrier view cursor to the left. Only valid while viewing carrier.

CONTROL Z Rapid movement of carrier view cursor to left. Only valid while viewing carrier square wave.

X Move carrier view cursor to the right. Only valid while viewing carrier square wave.

CONTROL X Rapid movement of carrier view cursor to right. Only valid while viewing carrier square wave.

SPACE Changes the state of the carrier square wave above the cursor. Only valid while viewing carrier square wave.

C Plays the pitch period at the Edit Cursor.

D Change display mode of active buffer. Each time you press, display alternates from pitch only to amplitude only to pitch and amplitude. Current display moded is indicated in the text window.

CONTROL D Delete the pitch period at the Edit Cursor. All speech to the right of the Edit Cursor is shifted to the left to fill in the space left after the deletion. Valid only on bottom buffer.

CONTROL E Erase contents of bottom buffer. If selected, the program will verify this command.

H Rapid movement of the Edit Cursor to the left in the active buffer.

J Normal movement of the Edit Cursor to the left in the active buffer.

K Normal movement of the Edit Cursor to the right in the active buffer.

L Rapid movement of the Edit Cursor to the right in the active buffer.

CONTROL H Rapid movement of the Left Cursor to the left in the active buffer.

CONTROL J Normal movement of the Left Cursor to the left in the active buffer.

CONTROL K Normal movement of the Left Cursor to the right in the active buffer.

CONTROL L Rapid movement of the Left Cursor to the right in the active buffer.

P Play back entire contents of selected speech buffer in the active buffer.

- O Play back contents of selected speech buffer between the cursors in the active buffer.
- I Increase or lengthen the pitch period at the Edit Cursor in the bottom buffer.
- M Decrease or shorten the pitch period at the Edit Cursor in the bottom buffer.
- CONTROL E Erase entire contents of the bottom buffer. You will be asked if you are sure you want to do this.
- CONTROL M (or N) Lower amplitude value at the Edit Cursor in the bottom buffer only. If using in a macro, you must use N and not CONTROL M as this is interpreted as a RETURN by the Apple II.
- CONTROL I Increase amplitude value at the Edit Cursor in the bottom buffer.
- F Sets the pre-filter value. Also sets range filter value. Default is 0 if female voice, 45 if male voice. Allowable range is 0 to 80. A value of 0 has no effect and a value of 80 has the maximum effect. The pre-filter, if non-zero, will filter all recorded pitch periods before they are placed into the top buffer, i.e. any pitch period that is smaller than the pre-filter value will be combined (added to) the next occurring pitch value. See CONTROL R for range filter.
- CONTROL F Combine one pitch period. What this does is to take the pitch period at the Edit Cursor and combine it with the next pitch period at the immediate right. This command only filters data in the top buffer. Its purpose is to insure that all pitch periods are good candidates for pasting to the bottom buffer.
- CONTROL R Range filter. Filters a range of pitch periods between the cursors. Filtering dependent upon value defined by "F" key (pre-filter value). Applicable only on data in top buffer.
- 1 to 9 Transfers the number of selected pitch periods at the top Edit Cursor and inserts them into the bottom buffer at the location of the bottom Edit Cursor. Functions independently of selected buffer.
- Q Quit the Edit mode and return to the First Level Menu. The contents of the top and bottom buffer are not affected unless you exit the First Level.

APPENDIX C: SECOND LEVEL (AMPLITUDE) EDITOR COMMAND SUMMARY

Note: All commands require that the CAPS LOCK key be pressed.

- ESC KEY List the Help Screen.
- B Blank speech data at Edit Bar. Changes the carrier data to zero so that only silence will be left. Also sets Sound Master volume level to zero. The "B" key is quite useful for separating phonemes. One example is the gap in "six" between "si" and "x". Another example is

a gap between "t" and "oo" in the word "two." Without this gap, the word would sound more like "do". Plosives such as "k" and "p" will stand out if followed by a short period of from one to two blank amplitude samples.

I or Up Arrow increases the amplitude level at the Edit Bar. Only audible with Sound Master.

M or Down Arrow lowers the amplitude level at the Edit Bar. Only audible with Sound Master.

J or Left Arrow scrolls all screen data to the left. Only works if speech if speech buffer is longer than 40 samples.

K or Right Arrow scrolls all screen data to the right. Only works if speech buffer is longer than 40 amplitude samples.

P Plays entire speech buffer.

O Plays speech buffer from left of screen to Edit Bar.

Z Remove all low frequency carrier data at Edit Bar. Acts as a high pass filter. Helps sharpen the sounds of plosives, such as "t" or "k". It can also improve the sound of Fricatives, such as "f" and "sh". Whenever you use the "Z" key, the amplitude value is displayed in reverse video.

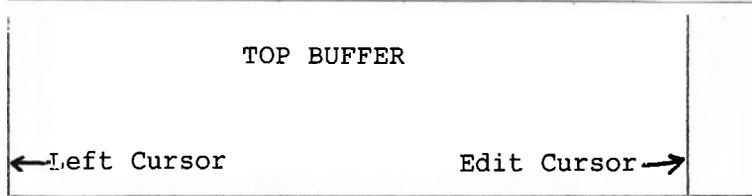
Remove every fourth cycle of carrier data at Edit Bar. Acts as a low pass filter. Each time you press X, the effect is exaggerated. The asterisk changes to 1, then 2, to a maximum of 3, depending upon how many times you press "X". It is possible to remove all but three periods of carrier if you continue to press "X".

Sibilent silencer. Asterisk, at Edit Bar, changes to "S" and is set to a volume level of 7. Instructs the speech playback software to lower the apparant volume with sibilents such as "ss" and "sh". Audible effect with Sound Master is reduced loudness level.

CONTROL S toggles Sound Master playback on or off. Not applicable with Ilc.

R Restore sample value. The "R" key restores any sample at the Edit Bar to the condition it had when you first entered the edit mode. If you leave the edit mode (by pressing "Q"), then any changes that you made become permanent.

Quit the editor and return to the Second Level (Amplitude) Main Menu. When you exit the editor, any changes to the amplitude sample values become permanent.



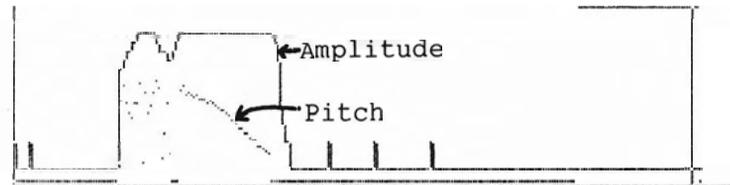
BOTTOM BUFFER

```

TOP ACCUM=00000      BOTTOM ACCUM=00000
DISPLAY=PITCH        BUFFER=TOP
SOUNDMASTER=ON      MACRO=
AMPLITUDES=00/00     PITCHES=000/000

```

FIG 1. - First Level Edit Screen



```

TOP ACCUM=32571      BOTTOM ACCUM=00000
DISPLAY=PITCH/AMP    BUFFER=TOP
SOUNDMASTER=ON      MACRO=
AMPLITUDES=03/00     PITCHES=189/000

```

FIG 2. - Recording of "saw" showing both Amplitude and Pitch displays.

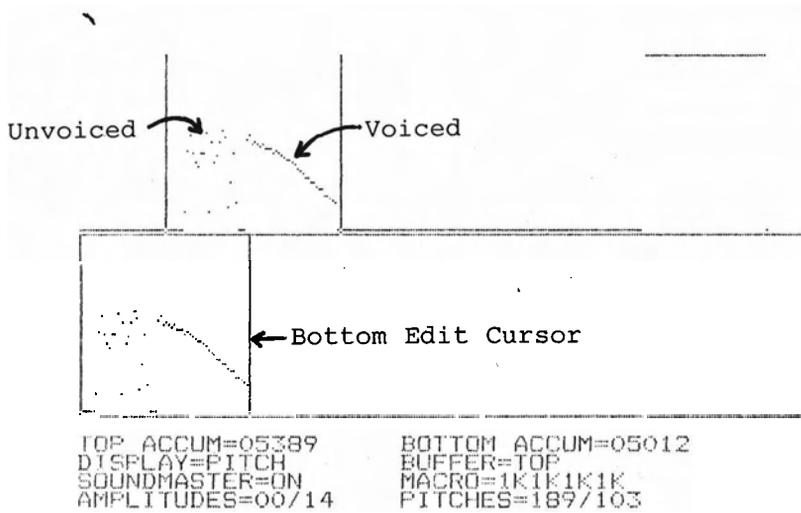
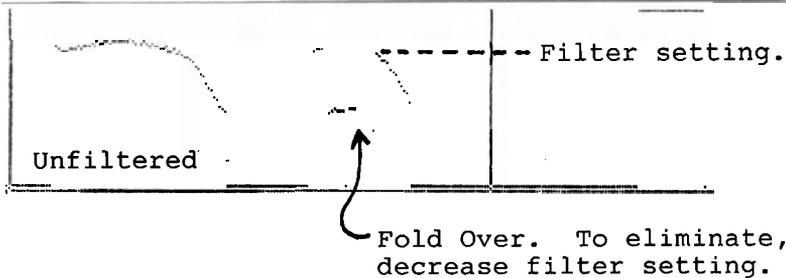


FIG 3. - Direct transfer of "saw" from the top buffer to the bottom.

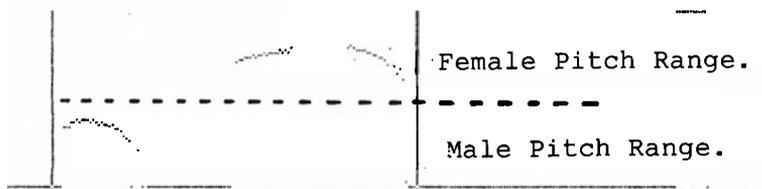


```

TOP ACCUM=18261      BOTTOM ACCUM=00000
DISPLAY=PITCH       BUFFER=TOP
SOUNDMASTER=ON     MACRO=1K1K1K1K
AMPLITUDES=00/00   PITCHES=189/189

```

FIG 4. - Two identical words spoken by a female. Second word exhibits "fold over."

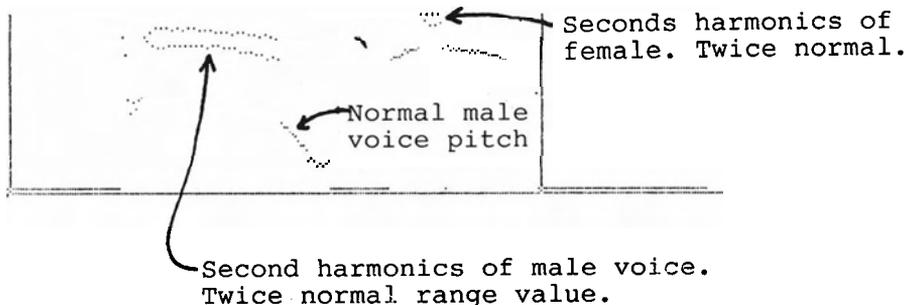


```

TOP ACCUM=11609      BOTTOM ACCUM=00000
DISPLAY=PITCH       BUFFER=TOP
SOUNDMASTER=ON     MACRO=1K1K1K1K
AMPLITUDES=00/03   PITCHES=189/189

```

FIG 5. - Same word spoken by male and female. Word is same length but higher pitches require more horizontal distance.



```

TOP ACCUM=16803      BOTTOM ACCUM=00000
DISPLAY=PITCH       BUFFER=TOP
SOUNDMASTER=ON     MACRO=1K1K1K1K
AMPLITUDES=00/10   PITCHES=189/126

```

FIG 6. - Second harmonic distortion. Two utterances of same word; first male, second female.



Pitch period examined
with View Option.

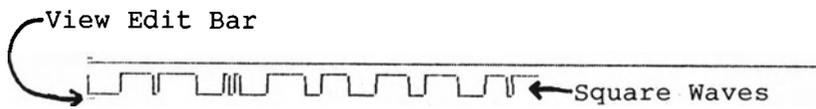
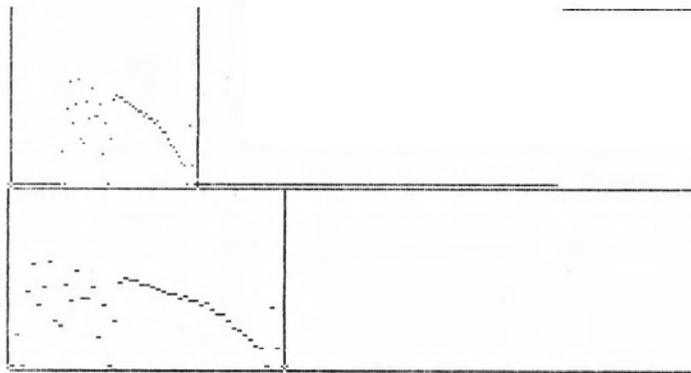
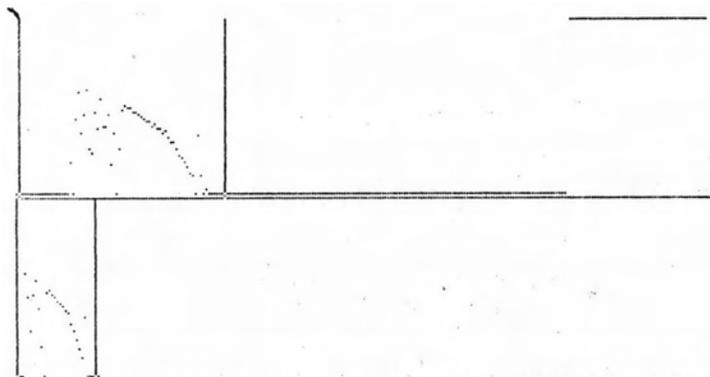


FIG 7. - View Display. Two words in top buffer same as Fig 6 except filtered using CONTROL-F.



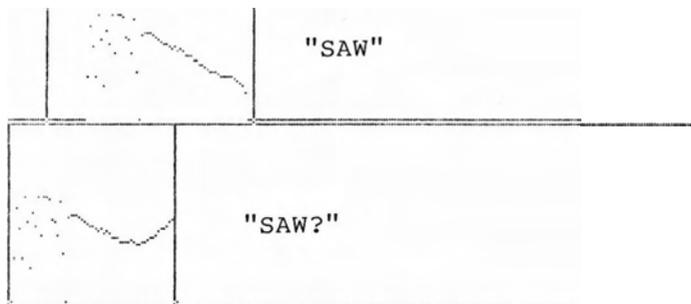
TOP ACCUM=07268	BOTTOM ACCUM=07733
DISPLAY=PITCH	BUFFER=BOTTOM
SOUNDMASTER=ON	MACRO=2K2K2K
AMPLITUDES=00/11	PITCHES=189/188

FIG 8. - Time expansion. Bottom word twice as long as top.



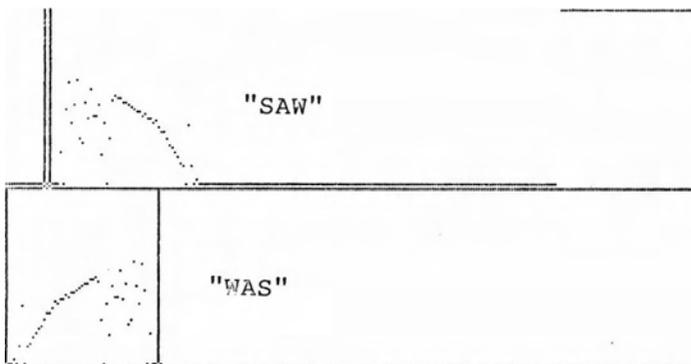
TOP ACCUM=08493	BOTTOM ACCUM=02653
DISPLAY=PITCH	BUFFER=BOTTOM
SOUNDMASTER=ON	MACRO=1KK1KK1KK
AMPLITUDES=00/00	PITCHES=189/189

FIG 9. - Time compression. Bottom word half as long as top.



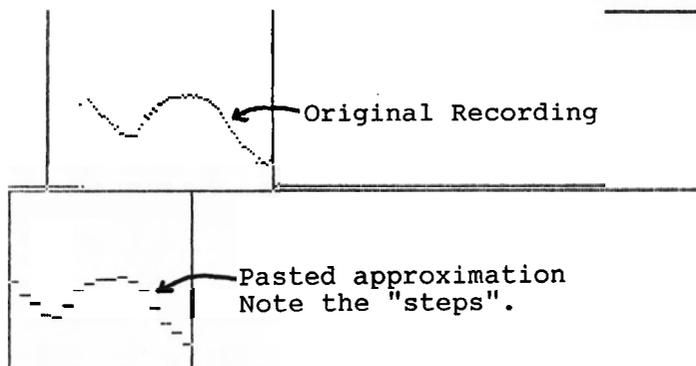
TOP ACCUM=07021 BOTTOM ACCUM=04105
 DISPLAY=PITCH BUFFER=BOTTOM
 SOUNDMASTER=ON MACRO=IIIIJ
 AMPLITUDES=00/15 PITCHES=189/184

FIG 10 - Alterting falling pitch of top buffer to rising pitch in bottom buffer.



TOP ACCUM=00378 BOTTOM ACCUM=04713
 DISPLAY=PITCH BUFFER=BOTTOM
 SOUNDMASTER=ON MACRO=1J1J1J
 AMPLITUDES=03/00 PITCHES=189/189

FIG 11 - Pasting top buffer backwards in bottom buffer.



TOP ACCUM=06864 BOTTOM ACCUM=04237
 DISPLAY=PITCH BUFFER=BOTTOM
 SOUNDMASTER=ON MACRO=4KKKK
 AMPLITUDES=12/12 PITCHES=055/083

FIG 12 - "Paste-and-repeat" of the word "away".

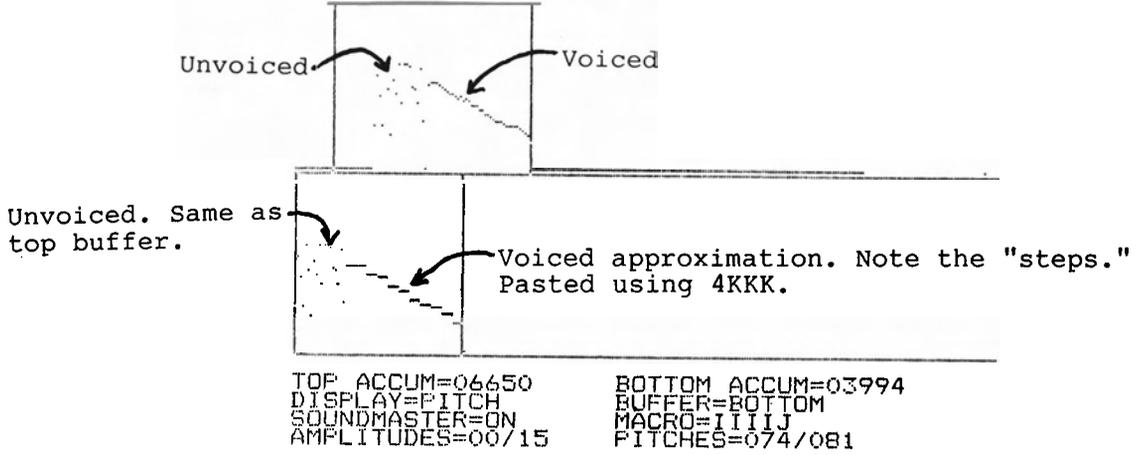


FIG 13 - Result of pasting "saw" from the top buffer to the bottom buffer using "4KKK" on voiced region only.

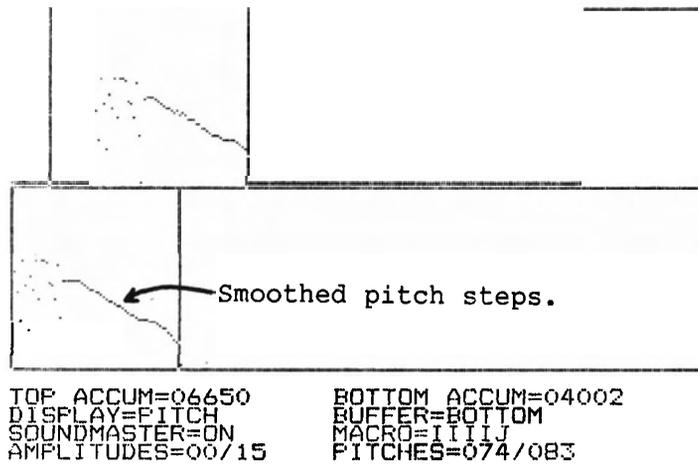


FIG 14 - Smoothing out of the "pitch steps" of Figure 13.

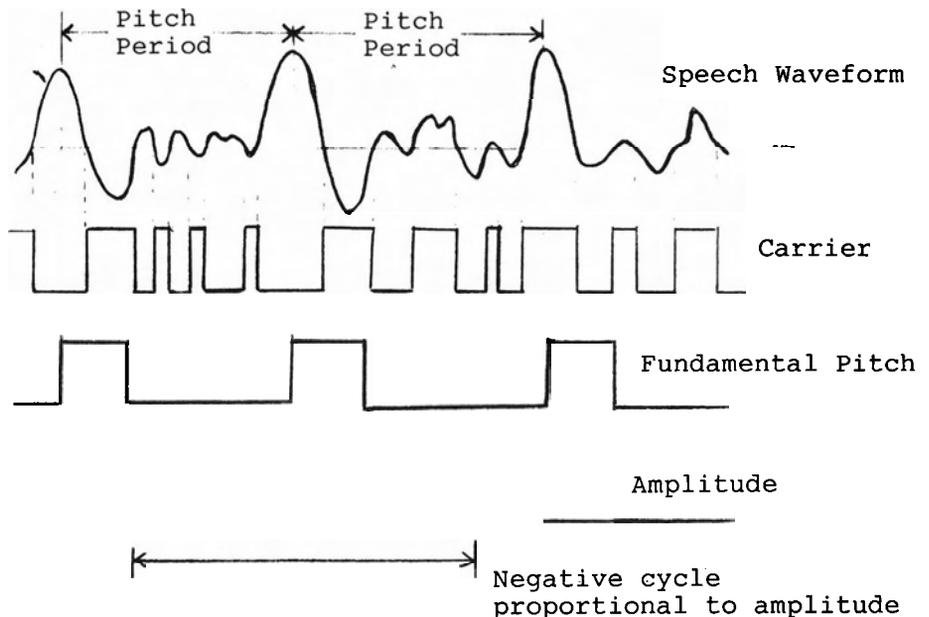


FIG 15 - Input waveform and the three waveforms produced by the Voice Master and analyzed by the software.



COVOX INC.

675-D Conger Street • Eugene, Oregon 97402 • U.S.A.
Area Code (503) 342-1271 • Telex 706017 (Av Alarm UD)