JANUARY 1984
HEWLETT-PACKARD JOURNAL

HEWLETT·PACKARD JOURNAL
Technical Information from the Laboratories of Hewlett-Packard Company

JANUARY 1984 Volume 35 • Number 1

Contents:

In this Issue:

Here in our *Hewlett-Packard Journal* offices, we recently took delivery on a new phototypesetting system. Our old system, acquired more than ten years ago, was a noisy nightmare of complicated electromechanical design. It had whirling drums, snapping relays, rotating mirrors, traveling lenses, and assorted shutters and rollers. Usually it worked, and it served us very well for many years. Our new system has only four moving assemblies, reflecting the universal trend towards doing more things electronically and fewer things mechanically, thereby gaining reliability. Significantly, two of our new typesetter's four moving mechanisms are disc drives—"floppies", like the ones you can see in nearly every personal computer system. From microfloppies to large multimegabyte drives, disc drives are an exception to the trend to more-electronic systems. Spinning platters coated with magnetic oxide are still the preferred media for storing massive amounts of data so a computer can have rapid access to any piece of it. Bubble memory, the only serious contender for the disc drive's role, has so far made inroads only where severe environments would knock out a disc drive's finely tuned mechanisms. A cost-saving feature of some disc drive designs is media removability, which lets users access more data with each drive by changing discs. It's taken for granted in a floppy drive, but becomes less cost-effective as drive capacity increases.

Our cover subject this month is the industry's largest removable disc drive media module. Its seven discs hold 404 megabytes, or over four hundred million characters, roughly the amount of information in all of the books in the photograph. It works in the HP 7935 Disc Drive, which is also distinguished by an unusually comprehensive set of built-in diagnostics. On pages 3 through 26, you can read about the design of this high-capacity drive and its companion fixed-media drive, the HP 7933.

Two speech output modules for HP computers are described in the articles on pages 29 and 34, one for HP Series 80 Personal Computers, and one for HP 1000 and HP 3000 Computers. The latter is compatible with the industrywide RS-232-C/V.24 interconnection protocol, so it can also be used with other manufacturers' computers if you can get along without the special software that makes the modules easier to use with HP computers. If you can't always be looking at the computer, it does help if the computer can talk.

-R.P. Dolan

Speech Output for HP Series 80 Personal Computers

This module allows a computer to provide informative prompts and alarms, freeing the user from frequent attention to a display.

by Loren M. Koehler and Timothy C. Mackey

THE HP 82967A SPEECH SYNTHESIS MODULE (Fig. 1) for Hewlett-Packard's Series 80 Personal Computers allows these machines to output audible information to prompt operators, sound alarms, indicate error conditions, or request service. The 82967A is accompanied by enough vocabulary and software to provide a variety of tools for using speech. A similar module, the HP 27201A, is available for computers with an RS-232-C/V.24 interface. A powerful supporting software package makes this module easy to use on the HP 1000 and HP 3000 Computer Systems (see article on page 34).

Computer-generated speech output can improve the efficiency of some operations, and in many cases, provide new capabilities. Some possible applications are:

- Test/measurement. A typical test station might use a voltmeter controlled by a Series 80 Computer via the HP-IB (IEEE 488). With speech output, the computer can tell a technician when to move a test probe to a new circuit node without the technician's having to divert attention away from the circuit board to obtain the next test instruction. The computer also can be programmed to warn the technician verbally when a high voltage is encountered.
- Data entry. An accounting clerk who enters data into ledgers all day long gets very good at using the 10-key numeric pad on the computer's keyboard. This user typically does not look at the screen during every data entry. Accuracy can be increased greatly by either having audible feedback of entered data or verbal warnings of "out-of-balance" account situations to such users, thus getting their attention and causing them to check their entries.
- Monitoring/process control. With speech output, a computer monitoring peripheral status or controlling a process can provide information with an audible alarm, eliminating the time required for an operator to locate the appropriate display and read the alarm message. For

Fig. 1. *The HP Model 82967A Speech Synthesis Module provides plug-in speech output capability for HP Series 80 Computers. The speech output can be supplied to headphones, an external audio system, or the speaker contained in the video monitors for HP-86 Computers. Included with the 82967A is a disc containing a 1500-word speech vocabulary and binary routines to simplify the application of speech output to programs.*

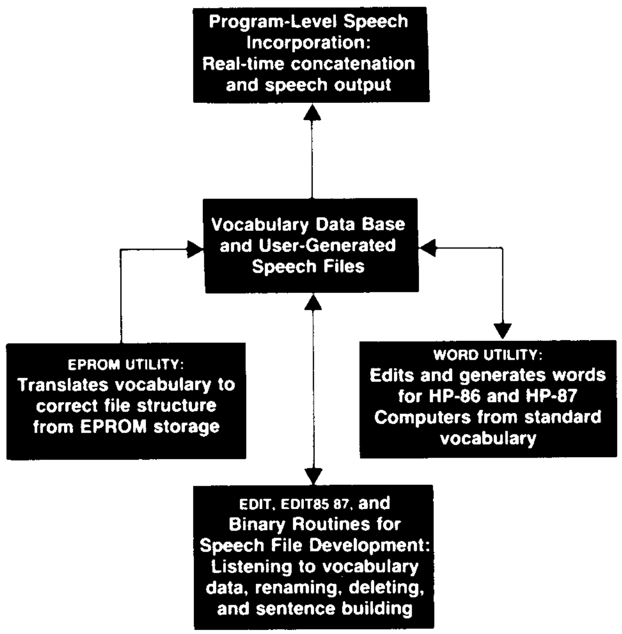

Fig. 2. *Outline of software package for implementing speech output in Series 80 Computers. All portions are included with the 82967A except the EPROM UTILITY and the WORD UTILITY, which are available from the HP User's Library.*

example, a computer can announce, "Printer 6 is out of paper," or "Furnace temperature is too high," to indicate not only the nature of the problem, but also the device affected.

Speech Output Technology

A computer system can use several techniques to generate speech output.[1] Some of the most common methods are converting text to speech, reconstructing speech using digitized samples or recordings of actual speech, and synthesizing speech using linear predictive coding (LPC, see box on page 32). The LPC technique is used by the 82967A and its companion product, the 27201A Speech Output Module.

Text-to-Speech Conversion. Translating text in ASCII characters into verbal output is done by examining the text in sequence and trying to figure out how to pronounce the syllables observed. Because most languages, in particular English, contain words or combinations of letters that look the same but sound different, or sound the same but look different, the set of translation rules can become quite complicated. For example, consider the italicized words in the following sentences:

Your assignment is to *read* the chapter about *reed* boats.
After you have *read* the instruction, pick up the *red* envelope.
After the archers picked up their *bows*, they faced the spectators and *bowed*.

To handle these problems, a large exception dictionary and an intricate set of context-dependent rules are required.

The advantage of translating text is its flexibility. However, the quality of the speech is generally very poor. The listener must pay close attention, because the speech output is usually flat and unemotional, and sounds mechanical. In some cases, the listener must have some knowledge of the general context of the spoken message to understand it correctly. Some of these deficiencies can be corrected, but then the benefit of flexibility is lost.

Recorded Speech. Actual speech can be digitized and stored in the computer's memory for reproduction when desired. The benefit is natural-sounding speech, but the memory requirements are large even for brief speech outputs, typically 12K bytes for speech lasting one second. An alternative approach is to use the computer to search for a desired phrase recorded on a tape recorder and direct the recorder to play back the phrase. The disadvantage of this method is the long access time and the large number of phrases required to verbalize a full range of numbers or conditions.

Synthesized Speech. Speech synthesis can generate high-quality speech using a mathematical model of the human vocal tract and digital data based on the recording of spoken words and sounds. Synthesized speech output has some clear advantages over more conventional methods of sound reproduction such as the tape recorder. Synthesized speech has solid-state reliability, allows real-time random selection and concatenation of words, phrases, and sounds, and has significantly lower data requirements than digitized speech. The LPC technique used in the 82967A and 27201A Modules is based on the assumption that the sound a person makes at one instant is a continuation of the sound made in the previous instant. LPC removes natural redundancies in speech and reduces the number of bits required to reproduce one second of speech by as much as 98.5% compared to purely digitized speech. Hence, the benefit of the LPC technique is that it greatly reduces memory requirements

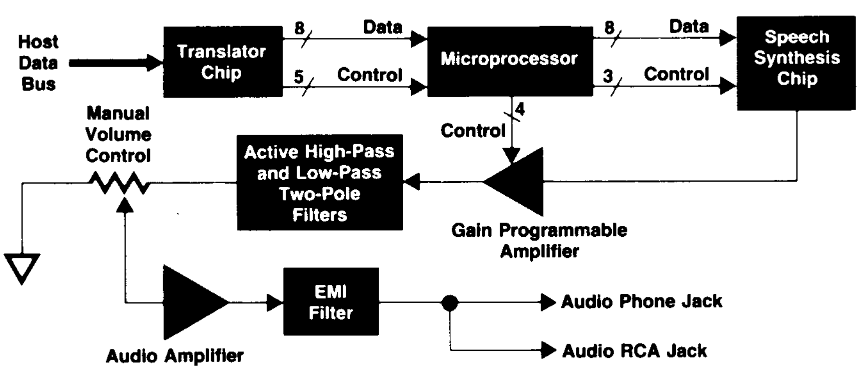

Fig. 3. *Block diagram of the 82967A Speech Synthesis Module's hardware system.*

while retaining a high-quality speech output.

System Overview

The design objectives for the 82967A were to provide speech output with as high a degree of quality as possible, and to provide everything needed to allow its use as a viable computing tool. The 82967A hardware is closely knit with software and vocabulary (refer to Fig. 2).

The EDIT program generates speech files. A speech file is a subset of the vocabulary data base, with words, phrases, and sounds put into user-defined combinations and renamed to fit the given application. A speech data file, as with any data file, requires a certain amount of time to load from mass storage into the computer and requires a certain amount of memory space. Speech file construction allows the user to make tradeoffs between vocabulary residing in computer memory and the amount of load time from mass storage.

The EPROM UTILITY program converts additional vocabulary words, usually residing on an EPROM (electrically programmable read-only memory), to the correct data file structure. EPROMs programmed with additional vocabularies selected by the user are available from the vendor for the speech synthesis integrated circuit used in the 82967A Speech Module. The EPROM UTILITY program is available from the HP User's Library service. It works in conjunction with the HP 82929A Programmable ROM Module for Series 80 Computers. The 82929A holds up to two 64K-bit EPROMs and fits directly into the Series 80 I/O backplane.

The WORD UTILITY program can be used to develop new words from the existing vocabulary. New words are constructed by deriving the required syllables and phonetic sounds from existing words in the vocabulary data base and patching them together. The WORD UTILITY program also allows a word, phrase, or sound to be altered by varying the digital parameters that define it (refer to the box on page 32).

Hardware Design

The 82967A hardware system (Fig. 3) is centered around an 8049 microprocessor and a TMS5220 speech synthesis chip. In addition, the module contains a programmable amplifier for attenuation of the volume under software control, active low-pass and high-pass filters, and a final audio amplifier.

The 8049 microprocessor services the speech chip, controls the gain programmable amplifier, and handles the I/O protocol between the 82967A and the host computer via a translator chip.

The IMB5 translator chip provides the interface between the module and the bus to the host computer. To service the speech chip, the microprocessor provides an internal 90-byte first-in, first-out (FIFO) buffer between the host computer and the speech synthesis chip. The computer invokes verbal execution of an utterance (a sound, a word, or a group of sounds and/or words) by passing to the host computer a command and a 16-bit value representing the total number of bytes in the speech file containing the data for the utterance. The computer then transmits the speech data for the utterance to the microprocessor's FIFO buffer. After it fills its FIFO buffer or receives all of the bytes in the utterance (if less then 90 bytes), the microprocessor sends a speak command and 16 bytes of speech data to the TMS5220 speech chip. The microprocessor then passes additional eight-byte blocks of speech data on a request basis to the speech chip until the utterance is finished. As data is moved from the microprocessor's FIFO buffer to the speech chip, more data is sent from the host computer until all of the requested speech data has been transferred from the computer to the speech module.

The data transfer used between the Series 80 Computer and the 82967A card is known as an OUTPUT data transfer. The OUTPUT data transfer was selected as the only means of transferring speech data from the computer to the speech module because this type of data transfer cannot be suspended by another device. Therefore, it can guarantee that all speech data in an utterance arrives at the 82967A without interruption. The OUTPUT data transfer approach reduces the memory requirements on the speech module, which keeps the hardware cost down. After completion of the OUTPUT speech data transfer, the other I/O operations of the computer can resume or begin while the speech module is still speaking, using the speech data remaining in the FIFO buffer.

The TMS5220 speech chip produces synthesized speech by taking the encoded LPC parameters, decoding the information, and placing it in an internal digital lattice filter. The ten most-significant bits of the filter output are supplied to an on-chip digital-to-analog converter (DAC) every 125 microseconds, yielding an analog signal representing the appropriate utterance at the output of the speech chip.

Programmable volume control is accomplished by using a gain programmable amplifier (GPA). This amplifier is actually a digital-to-analog converter. The analog signal from the speech chip is fed into the GPA's reference node and attenuated to one of sixteen possible levels, depending on the setting of the four binary inputs to the GPA by the microprocessor. The GPA setting is controlled from the keyboard or program by using the binary keyword SVOL, or alternately, the CONTROL command from the Series 80 I/O ROM. For example, the command SVOL 10,12 sets the output to level 12. The value 10 is the factory setting for the 82967A's select code (hardware address). Writing a value of 15 to the GPA sets the maximum volume level; a value of 1 yields a minimum volume level. Writing a value of 0 to the GPA turns off the output. At power-on or after a system reset, the speech module sets a default value of 15 for the GPA.

Two-pole high-pass and low-pass active filters are used to modify the analog signal before final amplification on the card. The −3-dB point of the high-pass filter is 100 Hz, and the −3-dB point of the low-pass filter is 3 kHz. These points were picked based on the bandwidth of human speech and background system noise. An operation called deemphasis, which attenuates the high frequencies of the synthesized speech, is performed directly by the TMS5220 speech synthesis chip. After the analog speech signal is filtered, it passes through a manual volume control and into an audio amplifier. The output of the audio amplifier is 0.22 watt into an eight-ohm load, which is more than

Linear Predictive Coding

Linear predictive coding (LPC) is a speech analysis/synthesis technique which reduces the amount of stored information required to reconstruct an utterance. If actual speech is digitized for storage, 96,000 bits of memory are required to reproduce one second of speech. By storing only the data required by LPC to synthesize the same one second of speech with high quality, the memory required is reduced to less than 1200 bits.

LPC analysis begins with a recording of the actual sound, word, or combination of words and/or sounds. This recording is then converted into digital data by first sampling the recorded waveform at a fixed rate. This data is then compressed to extract amplitude, source, and filter information to reconstruct the utterance based on a mathematical model of the human vocal tract. The amplitude, or energy, is simply the loudness of the utterance. The source information specifies whether or not the vocal cords are vibrating, and if so, at what frequency (pitch). The filter parameters describe the relative positioning of the tongue, lips, and teeth in the vocal tract model.

Speech Analysis

The filter parameters are derived based on the assumption that whatever is spoken now is to a large degree a linear continuation of what was spoken an instant earlier. That is, the filter parameters K_i are determined by minimizing the mean square error between the actual value v_n of a sample of the utterance and the value v_{ne} estimated from a weighted sum of a number of previous actual values. The relation between v_{ne}, the K_i coefficients, and the previous sample values is given by

$$v_{ne} = \sum_{i=1}^{j} K_i v_{n-i} \qquad (1)$$

where the analysis period, $i = 1$ to j, is a function of the sampling rate and the value chosen for j. For the LPC technique used in the 82967A and 27201A Modules, the sample rate is 8 kHz and $j = 10$.

Matrix algebra is used to solve for the K_i coefficients for each analysis period. The amplitude value is derived from the rms value of the speech waveform. The pitch value is derived from the periodic variation of the sample values.

Phonetics

Before discussing how the LPC parameters are combined to form a frame of speech data and how that frame is used to synthesize speech, the following information about basic phonetics will be useful.

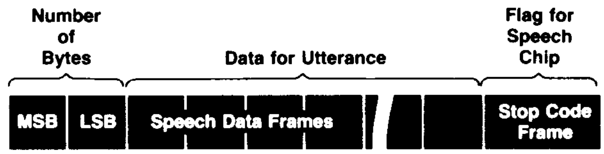

Fig. 2. *An utterance is coded as a sequence of speech frames with a header that indicates the total number of bytes in the sequence and a final stop speech frame.*

Speech sounds can be divided roughly into two categories: unvoiced and voiced. Unvoiced sounds are random noise generated by a constriction somewhere in the vocal tract. Examples of unvoiced sounds are the letters f and s. Voiced sounds involve the vocal cords vibrating at a certain pitch. All vowels are voiced sounds as well as the letters l, m, n, r, w, and y. Some letters such as v and z are combinations of unvoiced and voiced sounds.

Some letters have other characteristics that must be considered for speech synthesis. All of the consonant letters except h involve a constriction in the vocal tract that is relaxed rather quickly. For the stop letters b, d, g, k, p, and t, the constriction completely blocks the flow of air through the vocal tract. Hence, stop letters are characterized by a short silence lasting 10 to 100 milliseconds.

All letters characterized by either a partial or a complete vocal tract constriction are also characterized by rapid spectral (frequency) changes before and after the occurrence of the constriction. This is caused by the rapid changes in the structure of the vocal tract as the constriction is formed and then relaxed.

Speech Frames

The LPC parameters for each analysis period are stored with a repeat bit in speech frames of different lengths as shown in Fig. 1. Each word or sound is stored in the vocabulary as a sequence of these speech frames headed by a digital value that indicates the number of bytes in the sequence (Fig. 2).

The speech frames can be classified into five basic types: voiced, unvoiced, repeat, zero energy, and a stop code frame. The 50-bit voiced (V) frames specify energy, pitch, and all ten filter coefficients. The 29-bit unvoiced (U) frames specify energy, zero pitch, and only the first four filter coefficients. The 11-bit repeat (R) frames specify only energy and pitch, and have their

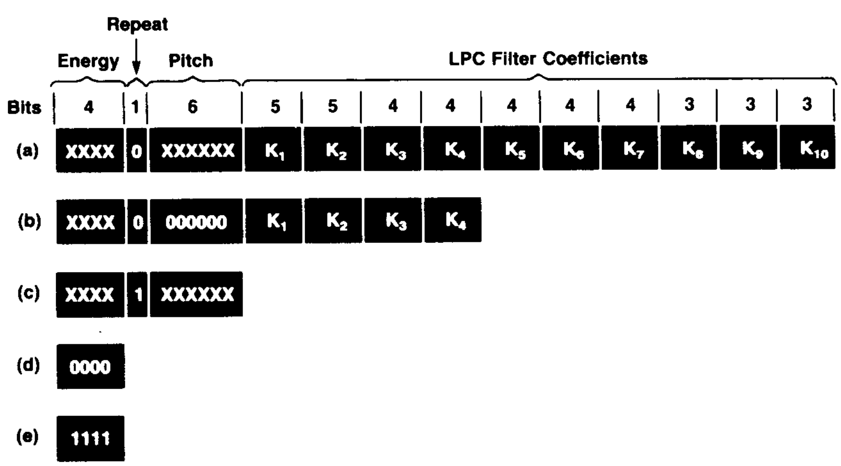

Fig. 1. *LPC speech synthesis parameters are stored into one of five speech frame formats: (a) voiced, (b) unvoiced, (c) repeat, (d) zero energy, and (e) stop frames (see text).*

Frame		Energy	Repeat	Pitch	K_1	K_2	K_3	K_4	K_5	K_6	K_7	K_8	K_9	K_{10}
V	-	12	0	43	22	11	6	3	10	11	10	6	3	3
V	-	13	0	39	25	13	9	2	4	11	8	6	4	4
V	-	13	0	38	25	13	9	2	3	11	9	6	4	4
V	-	13	0	38	24	16	8	1	4	11	9	6	3	3
V	-	13	0	41	25	17	8	0	3	11	9	6	3	4
V	-	11	0	45	24	14	6	0	5	13	10	6	3	2
V	-	7	0	49	21	12	2	3	12	13	12	6	2	3
V	-	1	0	1	25	8	10	5	5	4	2	5	3	0
Z	-	0	-	-	-	-	-	-	-	-	-	-	-	-
Z	-	0	-	-	-	-	-	-	-	-	-	-	-	-
Z	-	0	-	-	-	-	-	-	-	-	-	-	-	-
Z	-	0	-	-	-	-	-	-	-	-	-	-	-	-
U	-	1	0	0	21	6	6	4	-	-	-	-	-	-
U	-	1	0	0	24	0	6	8	-	-	-	-	-	-
U	-	7	0	0	31	15	1	2	-	-	-	-	-	-
U	-	7	0	0	27	1	3	9	-	-	-	-	-	-
U	-	7	0	0	31	10	4	5	-	-	-	-	-	-
UR	-	6	1	0	-	-	-	-	-	-	-	-	-	-
UR	-	4	1	0	-	-	-	-	-	-	-	-	-	-
S	-	15	-	-	-	-	-	-	-	-	-	-	-	-

Fig. 3. *Listing of the twenty speech frames and LPC parameter values in the sequence for the word "eights." The utterance begins at the top of the list and concludes with the stop frame at the bottom.*

repeat bit set to one to indicate to the speech synthesis IC that the filter coefficients for the previous frame are to be retained. The 4-bit zero energy (Z) frames specify zero energy and have no other values. The 4-bit stop (S) frame specifies an energy value of 15 and has no other values. The shorter length of the unvoiced, repeat, and zero energy frames allow the data rate and storage requirements for speech synthesis to be reduced considerably.

In each frame, the energy value can range from 0 to 15, and the pitch value can range from 0 to 63. The repeat bit is either zero or one. Note in Fig. 1 that the number of bits allocated for each filter coefficient varies depending on the influence of that parameter on speech quality.

Fig. 3 shows a listing of the values in each of the 20 frames in the 74-byte sequence of values required to synthesize the word "eights." The Z frames describe the momentary silence before the "t" and the U and UR frames describe the "ts" sound.

Speech Synthesis

Fig. 4 shows a block diagram of the basic speech synthesis process. The excitation sources are a variable-frequency pulse generator for voiced frames and a random noise generator for unvoiced frames. The amplitude of the selected source is modulated by the desired amplitude or energy level and then is applied to the recursive filter, which is programmed by the filter coefficients for the speech frame. The transmission function of the filter is

$$T(z) = 1/(1-H(z)) \quad (2)$$

where

$$H(z) = \sum_{i=1}^{10} K_i z^{-i} \quad (3)$$

Therefore, if e_n is the nth sample of the selected excitation, the synthesized output sample v_n' is

$$v_n' = e_n + \sum_{i=1}^{10} K_i v_{n-i}' \quad (4)$$

This synthesized output is then sent to additional circuitry to be shaped by high-pass and low-pass filters and amplified for output as audible speech.

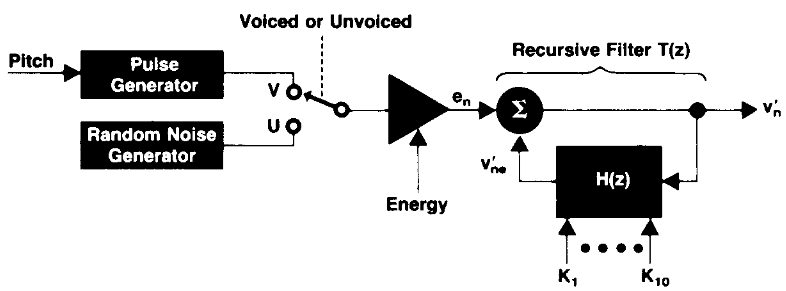

Fig. 4. *Block diagram of LPC speech synthesis process.*

adequate for most needs.

The 82967A comes with a phono jack and an RCA-type speaker jack connected in parallel. These two jacks enable the user to connect the speech module to an HP-86 Computer's video monitor (whose audio input also uses an RCA connector), to stereo headphones, or to whatever else meets the user's needs, without the need for an adapter. The manual volume control is designed so that by turning it all the way down, the correct volume setting is obtained for use with headphones.

High-Level Speech Incorporation

Incorporation of speech output in an HP BASIC program is a simple task. Special keywords provided by the "speak" binary program included with the 82967A's software allow quick implementation of speech output in programs. The software also contains easy-to-use routines for generating speech files that contain the desired utterances.

Retrieving Speech Data. To retrieve a speech file from mass storage, the procedure is the same as that required to read any data file into memory from mass storage with one exception—the keyword is different. For example,

```
10 ASSIGN #1 TO"HP" !Open the file
20 DLOAD;A$,P$ !A$ = name and file information, P$ = speech data
30 ASSIGN #1 * !Close the file
```

The above three HP BASIC statements load a speech file into computer memory. The actual speech data resides in the variable P$. Information related to the speech data such

Speech Output for HP 1000 and HP 3000 Computer Systems

by Elizabeth R. Hueftle and Jeffrey R. Murphy

The efficiency of some users of a large computer system can be notably improved by the addition of local speech output, which can be used to request service for peripherals such as plotters and printers, prompt operators for inputs, announce the occurrence of input or output error conditions, and sound alarms. The HP 27201A Speech Output Module in conjunction with the appropriate HP 27203A or HP 27205A Speech Library provides this capability for HP 1000 and HP 3000 Computer System users. The 27201A can also be used with other RS-232-C/V.24 computer systems, but the user will have to develop the speech files directly without the assistance of the Speech Library software.

The 27201A is a microprocessor-based peripheral whose appearance and internal speech synthesis circuitry are very similar to the 82967A Speech Synthesis Module for HP Series 80 Computers discussed in the accompanying article. The differences are the interfacing method and the addition of limited internal speech data storage.

Interface

The 27201A and its interconnecting cable implement a standard three-wire (for transmit data, receive data, and ground signals) RS-232-C/V.24 data communications interface. Because no other signals are provided, the module does not support connection to long-haul modems. However, short-haul modems using three-wire data communications can be used.

This interface scheme allows the 27201A to be connected directly to the host computer or to be inserted in series with any RS-232-C/V.24 peripheral such as a terminal, printer, or plotter. The serial configuration (Fig. 1) eliminates the need for another port when adding speech capability and allows the 27201A to be connected easily to a workstation for interactive speech output program development.

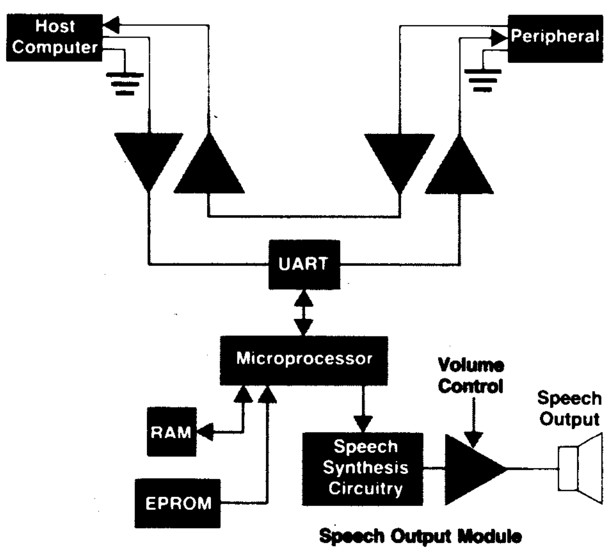

Fig. 1. *Serial configuration for HP 27201A Speech Output Module. The module is inserted in the RS-232-C/V.24 line connecting the peripheral to the host computer. During nonspeech operation, the module passes all signals between the peripheral and the host computer, but monitors the signals to detect the special escape command sequence that addresses the module.*

To coexist with an in-line peripheral on the same data communications line, the 27201A operates in an eavesdrop mode, monitoring all data on the line while passing it through to the in-line peripheral. The 27201A's speech output functions are invoked by a special escape sequence (Esc&yS<command>Esc&yU), which triggers command recognition by the speech synthesis circuitry. Table I lists the available commands. When the module must communicate with the host computer, it intercepts the peripheral's CTS (clear-to-send) line to suspend data flow from the peripheral during this time. The 27201A implements the XON/XOFF handshake, which means that data flow control resides with the 27201A, even when the peripheral initiates the handshake.

Table I

HP 27201A Speech Output Module Commands

Command	Action
CLEar	Deletes a group* of word data in the 27201A
DOWnload	Transfers word data from host computer to a group in the 27201A
PITch	Varies pitch of speech
RESet	Clears 27201A and runs self-test
SPEak	Causes 27201A to speak using word data in a group
STAtus	Identifies buffer, register, and memory status of 27201A and conveys error messages
TRAnsparent	Passes all data through 27201A without recognizing commands
UPLoad	Transfers word data from a group in the 27201A to the host computer

*A group in the 27201A represents one RAM or EPROM component.

Internal Memory

The Speech Output Module can store the data for synthesizing up to 100 words (50 seconds of speech) by downloading the data from the speech library in the host computer to the module's internal RAM, or can store the data for up to 200 words (100 seconds of speech) using EPROMs installed by the user. In the latter case, users can store the vocabulary most appropriate for their application and eliminate the time required to obtain speech data from the host system.

Speech Library Software

Fig. 2 illustrates the speech software package used by HP 1000 and HP 3000 Computers to generate speech output with the 27201A Speech Output Module. This package provides several useful features. First, it provides a vocabulary of over 1700 commonly used words and sounds in a form ready to be downloaded to a module and spoken. Second, the package provides easy access to these words and sounds. Third, the speech software provides a data base structure for the speech data and a program for managing that data base. Fourth, the software makes it easy for a user to include speech output in the user's application programs.

The VX (voice exerciser) program lets the user control the

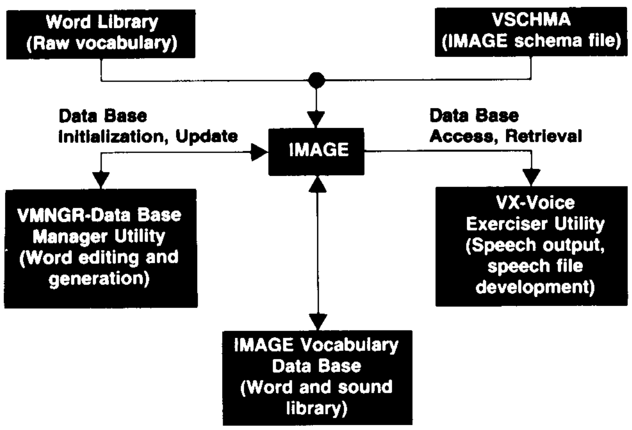

Fig. 2. *Outline of the Speech Library software package for HP 1000 and HP 3000 Computers.*

27201A without having to code module commands and speech data in a program. The interactive access provided by VX lets the user clear the module, download selected speech data to it, and cause it to speak by using only a few keystrokes. Thus, a user can listen to different outputs quickly to choose the best sounding one for the user's application.

The IMAGE schema file VSCHMA contains instructions used during the installation of the speech software to construct the data base, which is based on HP's IMAGE data base management system.

The VMNGR program manages the speech data base. It allows the user to:

- Install speech data into the data base
- Make other data bases out of subsets or supersets of the standard speech data set. For example, if an application requires a data base that takes up only a small amount of storage, VMNGR can be used to create a small data base containing only the words needed.
- Make up new words by combining parts of existing words. Instructions and hints about how to do this are included in the user manual.
- Add custom words to the data base. If a user needs words that are not in the standard vocabulary and cannot make them by modifying existing words, the user can purchase additional words from the vendor of the speech synthesis circuit used in the 27201A. These custom words are provided on an EPROM that the user can install in the speech module. VMNGR can then be used to access those words for installation in the central data base.
- Create a burn file for an EPROM. VMNGR can be used to format selected speech data into the proper file for programming an EPROM. The user can then program EPROMs for insertion in the memory sockets of the 27201A, which provides nonvolatile storage of speech data for fast speech output.

Acknowledgments

Many people contributed to the development of the speech products. Two of the individuals we would especially like to thank are Mike Thompson who developed the software for the HP 3000, and Dave Kenyon, our project manager, who was a source of continual encouragement and incredible enthusiasm. We would also like to thank the marketing team, especially Duncan Campbell and Tom Nakamura, for their contributions.

as the names of the words in the speech file and their location are contained in the variable A$. The binary keyword DLOAD provided by the binary speech software is used instead of the normal LOAD command because the vocabulary data base and user speech files have unique data structures.

Real-Time Concatenation. Combining words, phrases, and sounds residing in the host computer memory is an easy task for the applications program. One useful example is the generation of a number to be verbalized by a speech module in an HP-86 or HP-87 Computer. The statements

```
60 LET X$="20"
70 NUMBER$=PAR$("100",A$,P$)&PAR$(X$,A$,P$)
```

put the speech parameters for the word 120 in the string variable NUMBER$. This procedure enables any number between zero and 999 million to be constructed and verbalized in real time using the existing English vocabulary. To generate the given range of numbers, the speech file must contain speech data for the numbers zero through 20, the multiples of ten (30, 40, 50, etc.) through 100, and the numbers "thousand" and "million." The keyword PAR$( ) is part of the "speak" binary routine.

Speech Generation. To have a program speak an utterance is the most straightforward task of all. For example, to speak the number 120 generated above, requires only the program statement:

```
80 SPEAK 10; NUMBER$ !Where 10 is the 82967A's select code
```

A phrase of speech stored in a user's speech file can be as brief as one sound or word, or as long as thirty minutes of continuous speech. The phrases in a speech file can be named using a total of up to 195 characters on the HP-86/87 Computers and up to 95 characters on the HP-83/85 Computers by using the EDIT program. For example, in the statement

```
100 SPEAK 10; PAR$("DEMO",A$,P$)
```

DEMO is the label for the phrase "GOOD DAY PAUSE10 THIS IS THE HEWLETT PACKARD 8 2 9 6 7 A SPEECH SYNTHESIS MODULE." This phrase could have been constructed using the EDIT or the WORD UTILITY programs, or all the separate words could have been placed in separate phrases in a speech file and concatenated in real time to produce the same result.

Finding Words. The first step is to look up the desired words on the vocabulary disc that comes with the 82967A. This is done by running the EDIT program and pressing the VIEW softkey. The Series 80 host computer then asks the user which dictionary to look in. If the word wanted is "hello," the user enters H, the first letter in hello, because the standard vocabulary is arranged into 26 alphabetic dictionaries (one for each letter in the alphabet). A complete listing of the standard vocabulary on a removable placard is also included in the manual for the 82967A. User dictionaries are also allowed, and any name of up to eight characters can be assigned to a user-created dictionary.

Listening to Words or Phases. Once the user looks through the displayed words to confirm the existence of the desired words in the vocabulary, the user can listen to different sentence combinations by pressing the LISTEN softkey. The host computer then prompts the user by asking Word/Phrase to Speak? The user then enters each desired word separated by underscore marks. For example,

Word/Phrase to Speak?
HELLO_I_AM_THE_HEWLETT_PACKARD_EIGHTY_SIX_
PERSONAL_COMPUTER

The words are read from the vocabulary disc and the 82967A promptly "speaks" them. Next, a set of three softkeys (AGAIN, INCLUDE, CONTINUE) is displayed by the computer. The AGAIN softkey causes the phrase to be spoken again. The INCLUDE softkey causes the entire phrase to be brought into the dictionary residing in the host computer's memory. Pressing the CONTINUE softkey causes the original set of softkeys (VIEW, LISTEN, FET, SAVE, INCLUDE, DELET, RENAME) to be redisplayed.

Saving Results. Once the desired results are in the host computer's memory, the user can save this resident dictionary on disc by simply pressing the SAVE softkey. The computer then prompts the user for a name for the dictionary file this information will be stored in. Once the phrases are created and stored on disc, the user can write a BASIC application program to use them under program control with the help of the SPEAK command provided by the "speak" binary program that comes with the 82967A.

Packaging

The packaging scheme for the 82967A is designed so that the user does not have to assemble or disassemble anything to install the unit. The package is a plug-in module (Fig. 1), which in external appearance looks similar to the Series 80 HP 82950A Modem. A volume control and two audio jacks are the only clues that this is not an 82950A. Plugging the module into any of the four I/O slots in the Series 80 Computer backplane is the only installation needed (Fig. 4). Power is drawn from the mainframe, thus eliminating the need for any external power supplies and/or cords.

Fig. 4. *Installing the 82967A in a Series 80 Computer is a simple process. The module is plugged into one of the four I/O slots in the backplane of the computer as shown and derives its power from the mainframe.*

Acknowledgments

Homer Russell contributed to the definition of the speech system software and helped implement the 82967A for Series 80 Computers.

Reference

1. L. Rabiner and R.W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, 1978.

Bulk Rate

Hewlett-Packard Company, 3000 Hanover Street, Palo Alto, California 94304

HEWLETT·PACKARD JOURNAL

JANUARY 1984 Volume 35 • Number 1

Technical Information from the Laboratories of Hewlett-Packard Company
Hewlett-Packard Company, 3000 Hanover Street
Palo Alto, California 94304 U.S.A.
Hewlett-Packard Central Mailing Department
Van Heuven Goedhartlaan 121
1181 KK Amstelveen, The Netherlands
Yokogawa-Hewlett-Packard Ltd., Suginami-Ku Tokyo 168 Japan
Hewlett-Packard (Canada) Ltd.
6877 Goreway Drive, Mississauga, Ontario L4V 1M8 Canada

CHANGE OF ADDRESS: To change your address or delete your name from our mailing list please send us your old address label. Send changes to Hewlett-Packard Journal, 3000 Hanover Street, Palo Alto, California 94304 U.S.A. Allow 60 days.

5953-8519