Interfacing a single-board computer to a
digital voice recorder for use in developing a
speech recognition system

by

Gary A. Fagan

A Thesis Submitted to the

Graduate Faculty in Partial Fulfillment of the

Requirements for the Degree of

MASTER OF SCIENCE

Interdepartmental Program: Biomedical Engineering
Major: Biomedical Engineering

Iowa State University
Ames, Iowa

1990

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

DEDICATION

To the memory of my grandparents

INTRODUCTION

The purpose of this project was to develop a computer-controlled digital voice recorder by interfacing a digital voice recorder to a single-board computer for use in a biomedical application. The single-board computer is the MC-1Z by Basicon, Inc., and the digital voice recorder is the Digital Voice Module (DVM-1) by Ming Engineering & Products, Inc.

The MC-1Z is a low-cost, portable microcontroller that can be programmed in BASIC and/or machine language. Once the MC-1Z is supplied with +5 Vdc power and connected to an RS-232 compatible terminal, it is ready to be programmed for many control applications.

The DVM-1 is one of several digital voice recorders available on the market today. These recorders are similar in operation to audio cassette recorders except that they store voice data in memory integrated circuits (chips) instead of using magnetic tape. The advantages of memory chips are that they last longer than tape and allow more convenient access to voice data. The disadvantage is that the memory chips used in this project must be constantly powered to maintain storage of the voice data.

One possible application of the computer-controlled recorder is speech recognition. Since the MC-1Z computer would be a component of the speech recognition system, an

obvious application of the speech recognizer is verbal control of the MC-1Z. This capability would allow a person, unable to type on a keyboard, to operate and program the MC-1Z computer.

The implementation of the MC-1Z/DVM-1 speech recognition system is described in this thesis. The first section provides information on the components and operation of the DVM-1 recorder and the MC-1Z computer and explains the hardware and software used in interfacing the recorder and the computer. Additional hardware and software information is given Appendices A, B, and C.

The second section gives an overview of speech recognition including the problems, categories, processes, and applications of speech recognition. The remainder of this section is a literature review of two speech recognition systems that can be built by the experimenter. Finally, a discussion is given about the work done on the MC-1Z/DVM-1 speech recognition system and what needs to be done to enable this system to recognize speech. Appendix D is a listing of manufacturers of speech recognition systems.

A COMPUTER-CONTROLLED DIGITAL RECORDER

The Digital Voice Module

The main components of the DVM-1 are labeled in Figure 1.
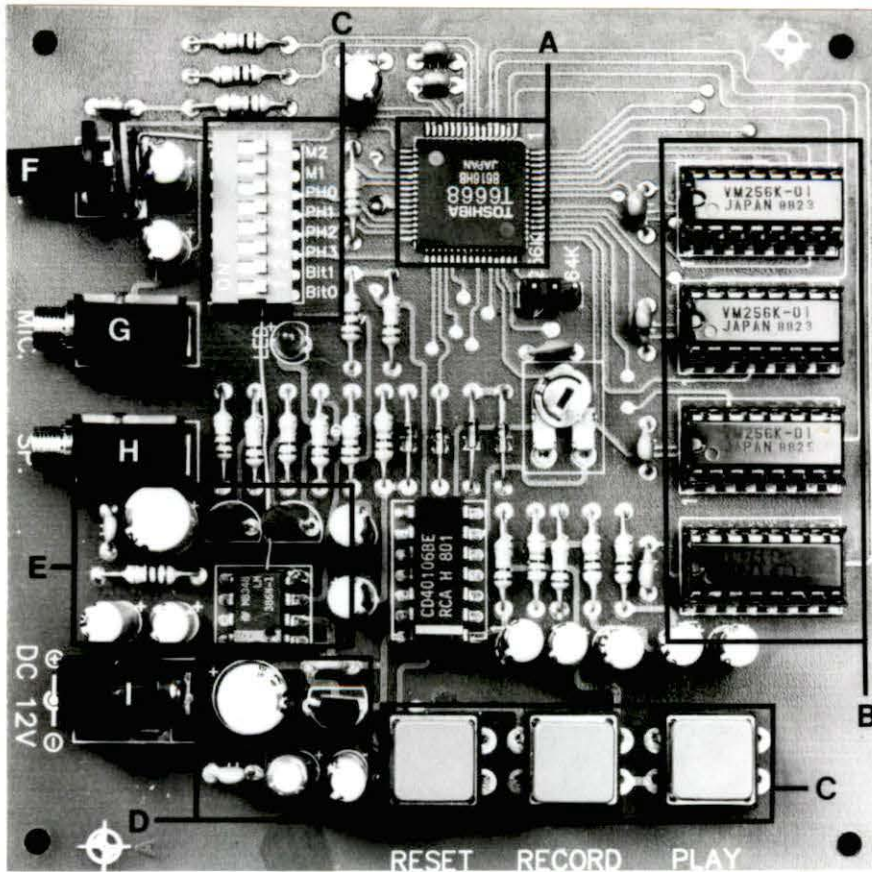
Figure 1.   Top view of the Digital Voice Module with the main components labeled: (A) T6668 speech microprocessor, (B) Memory, (C) Control switches, (D) Power circuit, (E) Audio circuit, (F) Volume control, (G) Microphone jack, (H) Speaker jack, and (I) Power jack

The T6668 microprocessor

The T6668 consists of several functional components shown in Figure 2. The pin connections are shown in Appendix A.
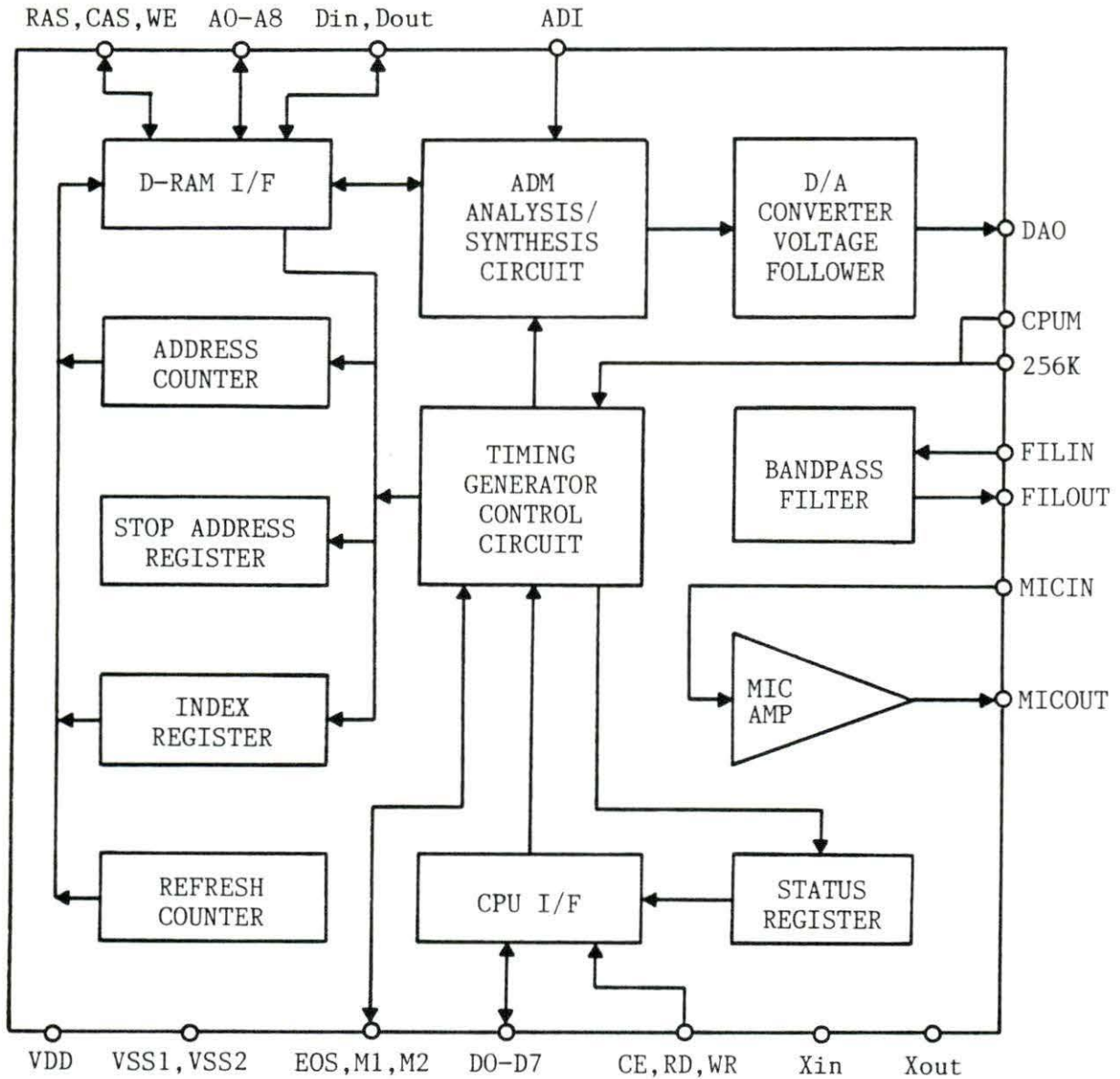
Figure 2. Block diagram of the T6668 showing the functional components (Toshiba America, Inc., 1988)

Microphone amplifier   The T6668 includes two
microphone amplifier circuits providing three possible
amplifier configurations. The amplifier circuits are shown in
Figure 3 and the three amplifier configurations are listed in
Table 1. The DVM-1 uses the third amplifier configuration.
Figure 4 shows the connections between the T6668 and the
microphone. The frequency response is flat between 100 and
10,000 Hz with a gain of 46 dB. The amplifier output (MICOUT)
is connected to pin ADI which is the input for the ADM
(adaptive delta modulation) analysis/synthesis circuit
discussed in the next section.
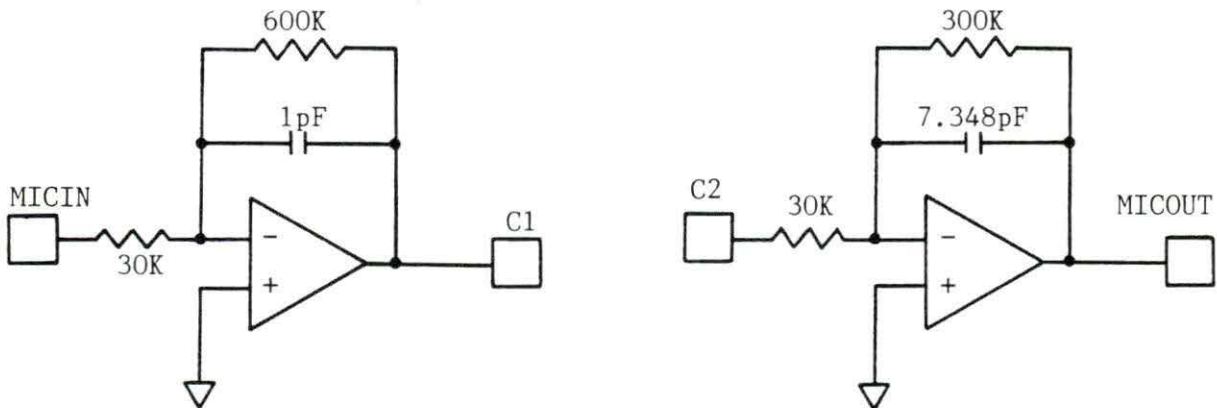
Figure 3.   The equivalent circuits of the two microphone
            amplifiers of the T6668 (Toshiba America, Inc.,
            1988)

ADM analysis/synthesis circuit   The ADM analysis/
synthesis circuit receives the amplified microphone signal at
pin ADI. The function of this circuit is to convert the
analog microphone signal into a digital signal that can be
stored in the D-RAM. Some of the methods of analog-to-digital

conversion include pulse code modulation (PCM), differential pulse code modulation (DPCM), and delta modulation (DM) which includes linear delta modulation (LDM) and adaptive delta modulation (ADM).

Table 1.   Three possible amplifier configurations using the two available T6668 microphone amplifiers

Configuration Number	Amplifier Input	Amplifier Output	Gain (dB)
1	MICIN	C1	26^a
2	C2	MICOUT	20^a
3	MICIN	MICOUT	$46^{a,b}$

[a]A 1 uF capacitor must be connected between the microphone and the input.

[b]Amplifier configuration 3 requires that pins C1 and C2 be coupled with a 1 uF capacitor.
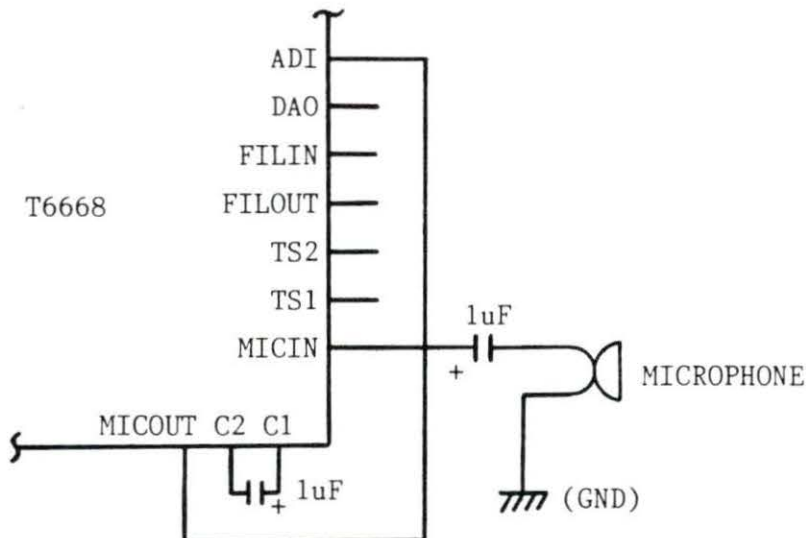
Figure 4.   Connections between the microphone and the T6668 to achieve the third amplifier configuration (Toshiba America, Inc., 1988)

Pulse code modulation (PCM)   Pulse code

modulation involves sampling, quantization, and binary

encoding of the input signal as illustrated in Figure 5.

	Sample value	1.3	3.6	2.3	0.7	−0.7	−2.4	−3.4
Nearest quantization level		1.5	3.5	2.5	0.5	−0.5	−2.5	−3.5
Code number		5	7	6	4	3	1	0
Binary representation		101	111	110	100	011	001	000
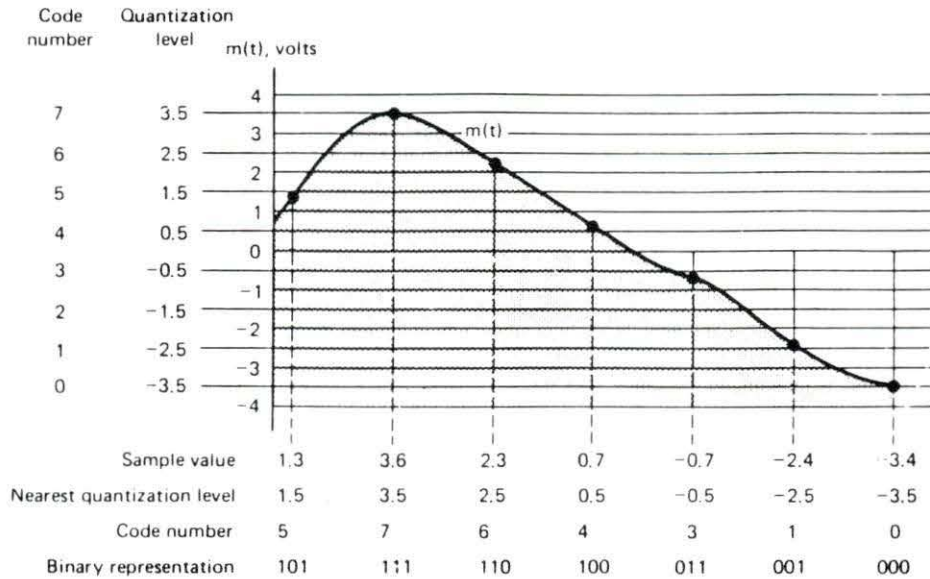
Figure 5.   An example of pulse code modulation showing
            sampling, quantization, and binary encoding (Feher,
            1987)

To accurately represent the analog signal, the sampling rate

should be at least twice the bandwidth of the input signal

according to Nyquist's sampling theorem. At each point of

sampling, the analog signal has a certain value called the

sample value. The sample value is equated to the nearest

quantization level. In this example, eight quantization

levels exist varying by one volt from −3.5 to +3.5 volts.

Each level is assigned a code number from 0 to 7. The binary

representation of the code number is the information used to

describe the analog signal.

<u>Differential pulse code modulation (DPCM)</u>   Differ-
ential pulse code modulation is a variation of PCM. Instead
of determining a quantization level for each sample value,
DPCM predicts a future sample value from the previous sample
value. The difference between the predicted value and the
actual value is called the prediction error. The prediction
error is quantized, coded, and sent to a decoder which
reconstructs the original signal (Feher, 1987).

<u>Delta modulation (DM)</u>   A version of DPCM, delta
modulation estimates the next sample value as a positive or
negative increment from the previous sample value. When the
increment or step size is constant (whether decreasing or
increasing), delta modulation is referred to as linear
(nonadaptive) delta modulation (LDM), and when the step size
can vary, it is called adaptive delta modulation (ADM).

In linear delta modulation (LDM), the previous sample
value is subtracted from the input signal. If the difference
is positive, the previous sample is increased by a step to
establish the next sample value, and if the difference is
negative, the previous sample value is decreased by a step
(Feher, 1987). LDM produces a signal with a maximum slope
equal to the step size times the sampling rate. When the
input signal exceeds this maximum slope, slope overload
distortion occurs. Another source of error is granular noise
caused by the difference in value between the actual signal

(X) and the LDM signal (Y). These sources of quantization noise are illustrated in Figure 6.
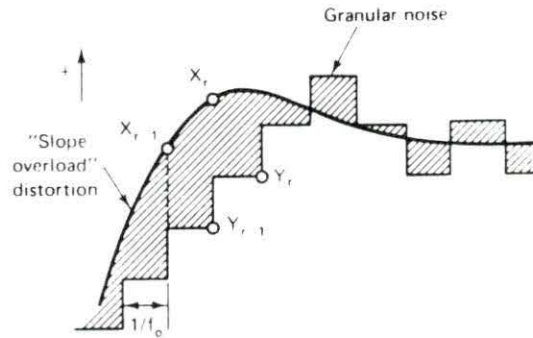
Figure 6.   Quantization noise in linear delta modulation
             (Feher, 1987)

In adaptive delta modulation (ADM), a digital algorithm is used to adapt the stepping process so that quantization noise is reduced as shown in Figure 7. Depending on the manufacturer, various algorithms are used to implement ADM, a system which provides better signal reproduction than linear delta modulation without increasing the sampling rate.
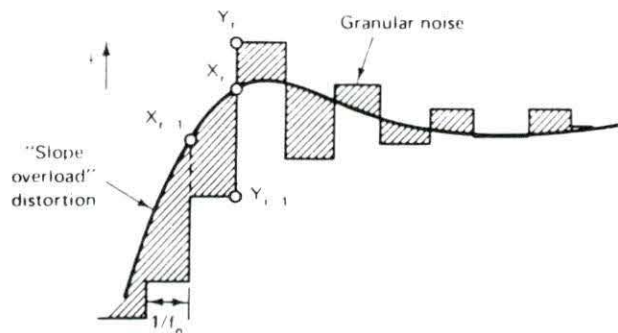
Figure 7.   Quantization noise in adaptive delta modulation
             (Feher, 1987)

D-RAM I/F   The D-RAM I/F component is an interface circuit that allows the T6668 to be connected to dynamic random access memories (D-RAMs). The T6668 accepts up to four pieces of D-RAM which may be either 64 Kbit or 256 Kbit D-RAM, but both types cannot be used together (Toshiba America, Inc., 1988). The pin connections between the T6668 and a D-RAM are shown in Figure 8. The functions of these pins will be explained in the section concerning memory.

Address counter   The address counter is a 20-bit counter that stores the address of the current memory location in D-RAM. The value of the counter can be set or read (Toshiba America, Inc., 1988).

Stop address register   The stop address register also has 20 bits and stores the address at which sound recording/ reproduction stops. Values may be written into the register, but they cannot be read (Toshiba America, Inc., 1988).

Index register   The index register specifies the end address of the index area. The index area in memory stores the start address, stop address, and bit rate of a recording (Toshiba America, Inc., 1988). The user does not have direct access to this register.

Refresh counter   Dynamic RAM stores information as charges on capacitors, and due to leakage, the capacitors must be recharged or refreshed to maintain the memory. The refresh counter is an 8-bit counter that refreshes 256 addresses of

T6668

CAS4

CAS3

NC

CAS2

CAS1

+5Vdc

RAS   DOUT   A1   VDD   A3   A5   A7
WE   DIN   A0   NC   A2   A4   A6   A8

A8   VSS

DIN   CAS

W   DOUT

RAS   A6

A0   A3

A2   A4

A1   A5

VDD   A7

256K D-RAM

0.1uF

GND
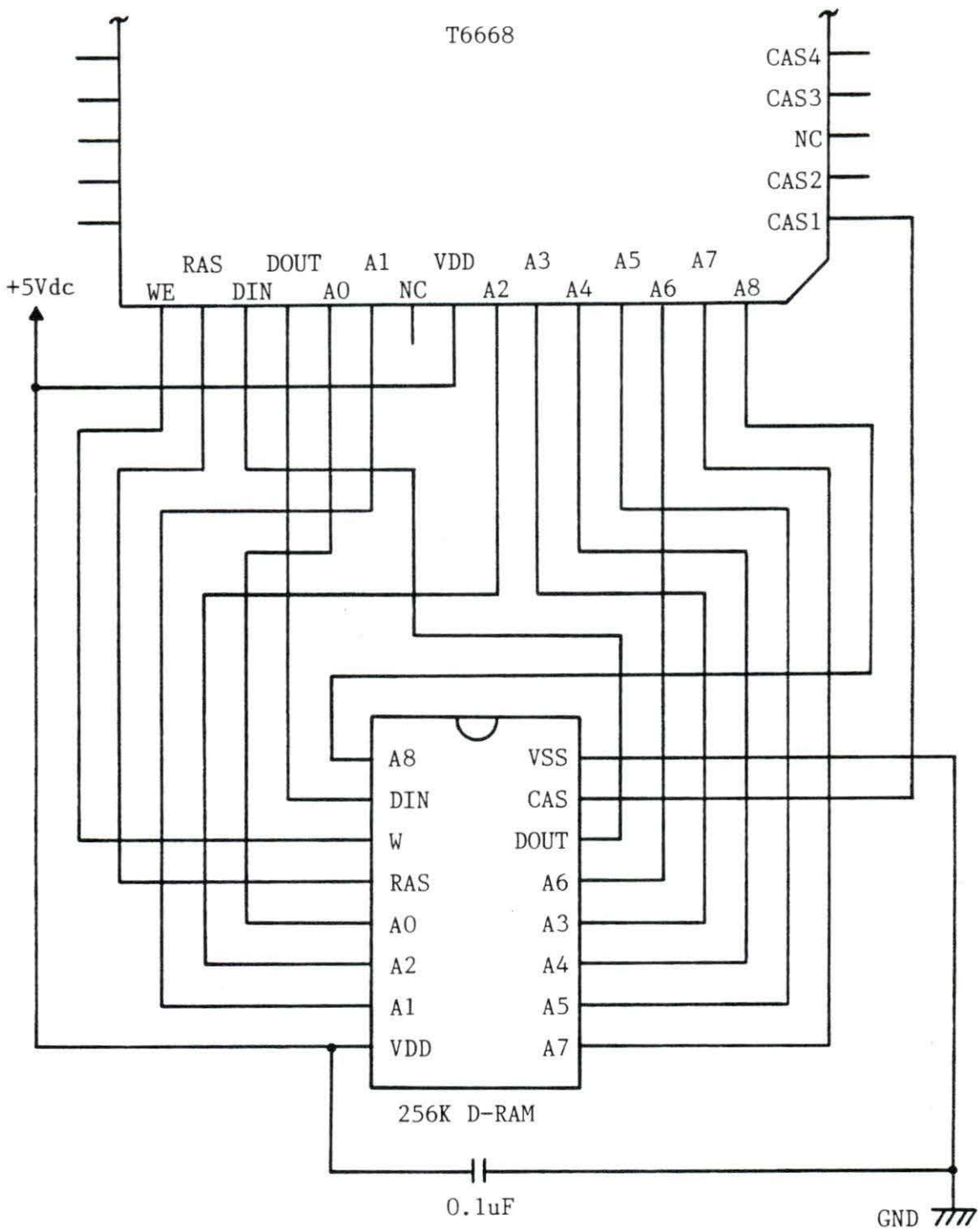
Figure 8.   Connection of D-RAM to the T6668 (Toshiba America,
            Inc., 1988)

the D-RAM within 4 milliseconds (Toshiba America, Inc., 1988).

Timing generator control circuit   The timing generator control circuit controls the operating rate of the ADM analysis/synthesis circuit, address counter, stop address register, and index register. Pins associated with the timing generator control circuit include EOS, M1, M2, 256K, Xin, Xout, and CPUM.

The EOS pin is low when recording or reproduction starts and high when recording or reproduction stops (Toshiba America, Inc., 1988).

Pins M1 and M2 are set according to the number of D-RAMS used as indicated in Table 2.

Table 2.   Logic levels of M1 and M2 based on the number of D-RAMS used (Toshiba America, Inc., 1988)

Number of D-RAMs	M2	M1
1	L	L
2	L	H
3	H	L
4	H	H

Note: H is +5 volts DC and L is ground (0 volts).

The 265K pin is set to logic low when using 64 Kbit D-RAM or logic high for 256 Kbit D-RAM (Toshiba America, Inc., 1988).

Xin and Xout are the input and output pins of the oscillator circuit operating at 655 KHz (Toshiba America, Inc., 1988).

The T6668 can be operated in a manual control mode or a CPU control mode. The DVM-1 uses manual control; therefore, CPUM is set low. For CPU control, the CPUM pin must be set high (Toshiba America, Inc., 1988).

D/A converter/voltage follower   During voice reproduction, voice data are retrieved from the D-RAM, processed by the ADM analysis/synthesis circuit, and delivered to the D/A converter/voltage follower circuit. This circuit converts the digital voice data into an analog signal made available at pin DAO. After bandpass filtering, the analog signal is sent to an audio circuit external to the T6668.

Bandpass filter   The T6668 contains a bandpass filter used during sound reproduction. The input of the bandpass filter is pin FILIN, and the output is pin FILOUT. The equivalent circuit of the filter is shown in Figure 9. The first stage is a high-pass filter, and the second stage is a low-pass filter.

CPU I/F   The CPU I/F is an interface circuit that allows the T6668 to be controlled by a central processor unit (CPU). The pins associated with this circuit include D0-D7, CE, WR, and RD. The functions of these pins depend on whether manual or CPU control is used.
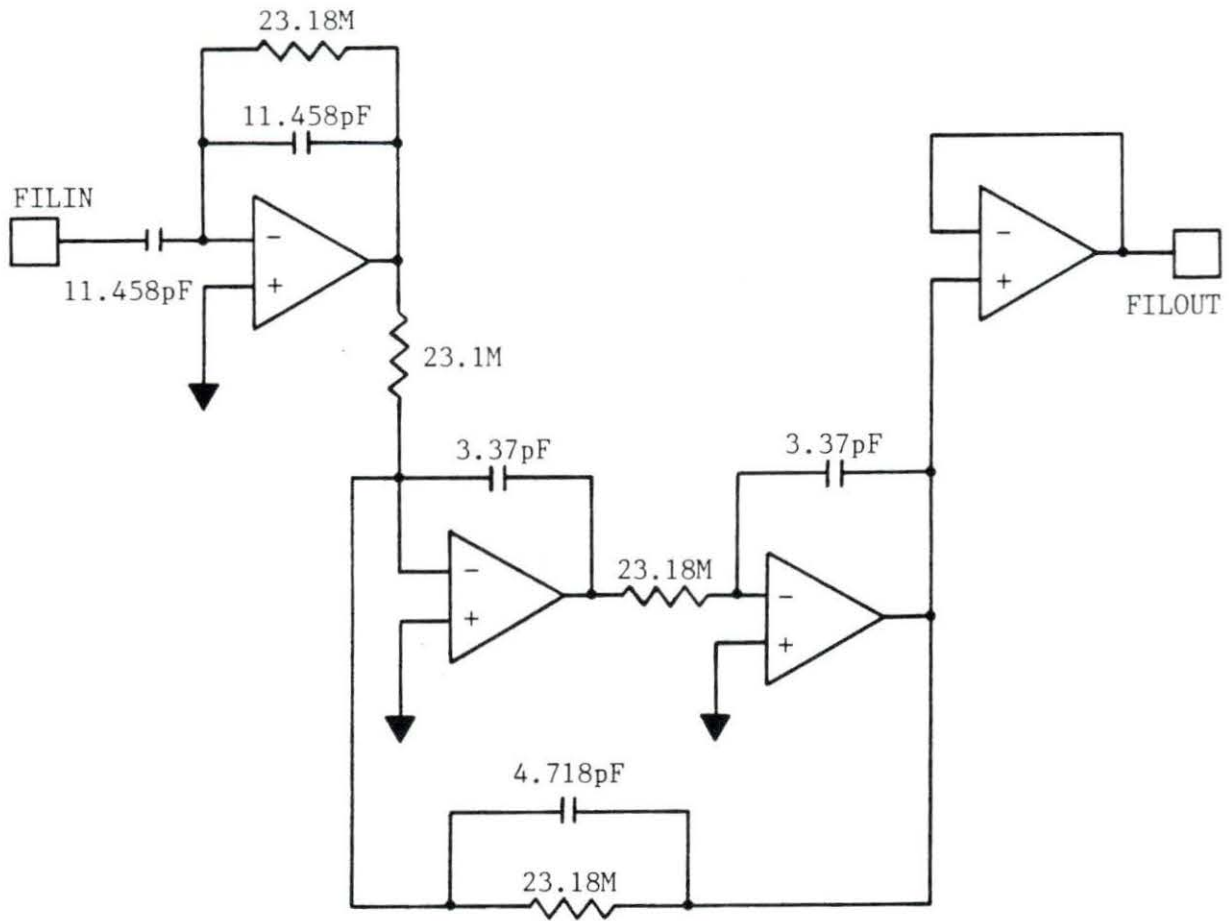
Figure 9.   Equivalent circuit of the bandpass filter (Toshiba
            America, Inc., 1988)

Manual control   Pins D0-D3 are the inputs for
selecting one of 16 independent phrases. Setting D4 high
starts voice recording or reproduction and setting D5 high
stops voice recording or reproduction (Toshiba America, Inc.,
1988). Pins D6 and D7 are inputs for selecting one of four

available bit rates. Pin CE must be set low to record and to allow voice output at pin DAO (connected to the input of the bandpass filter). Pin WR is set high to select the recording mode, and set low for the reproduction mode (Toshiba America, Inc., 1988). Pin RD is not used for manual mode. Connections for using the T6668 in the manual mode are shown in Figure 10.

CPU control   Pins D0 to D7 constitute a bi-directional data bus used for inputting commands to the T6668 or outputting the status of the T6668. Pins CE, RD, and WR are used to determine a write operation, read operation, or neither. When CE and RD are low, the status register can be read. When CE and WR are low, commands can be written. Connections for using the T6668 in the CPU mode are shown in Figure 11.

Status register   The status register is an eight-bit register that stores the status of the T6668. The status is read under CPU control by setting pins CE and RD low causing the status data to be output on pins D0 to D7. The designation of each status register bit is shown in Table 3.

When the busy bit is high, the T6668 is resetting itself or processing a command, and it should not be given commmands until the busy bit is low. Having the same value as pin EOS of the T6668, the EOS bit of the status register is low after recording/reproduction stops. The error bit (ERR) is high
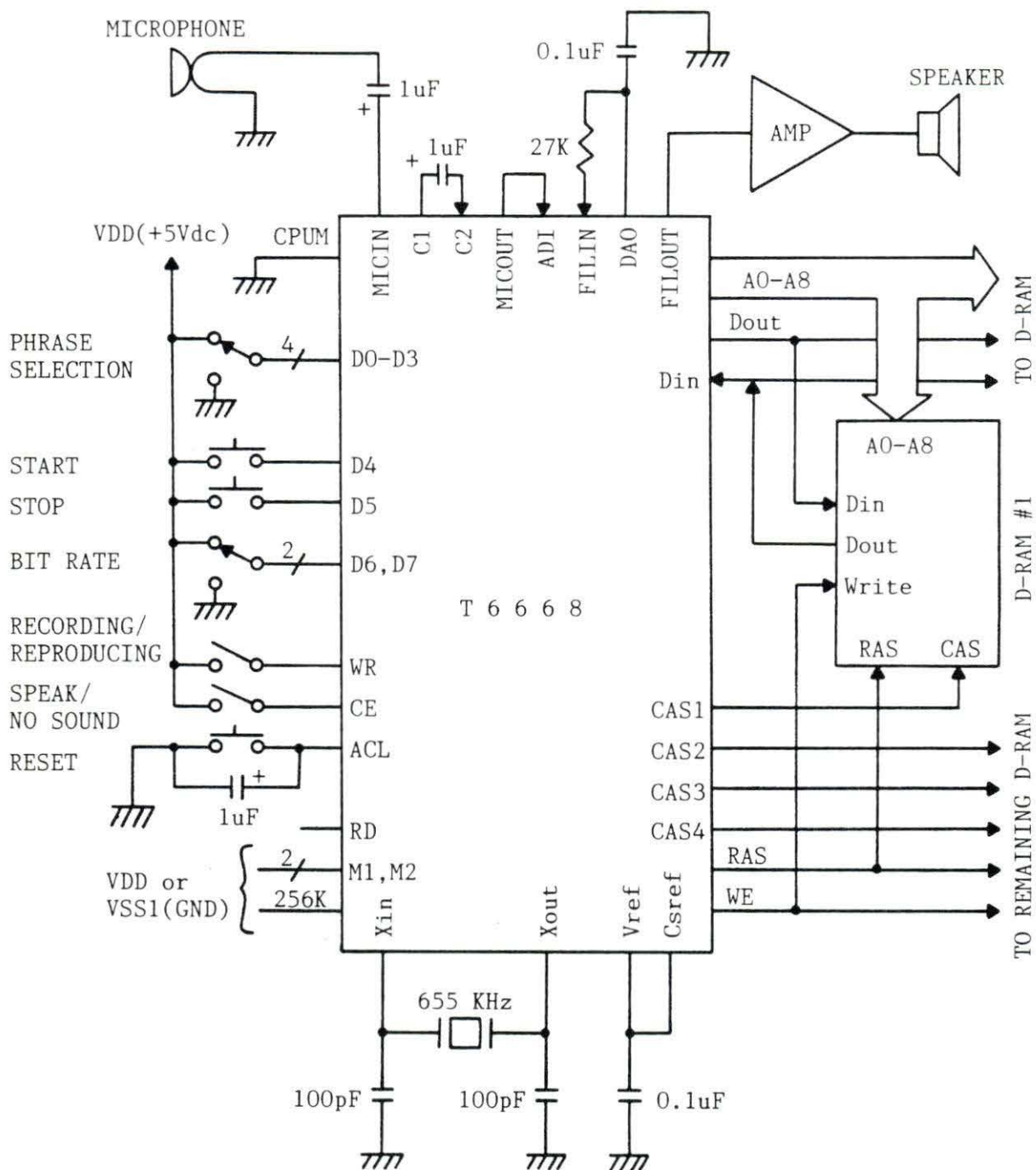
Figure 10. The T6668 wired for the manual mode (Toshiba America, Inc., 1988)
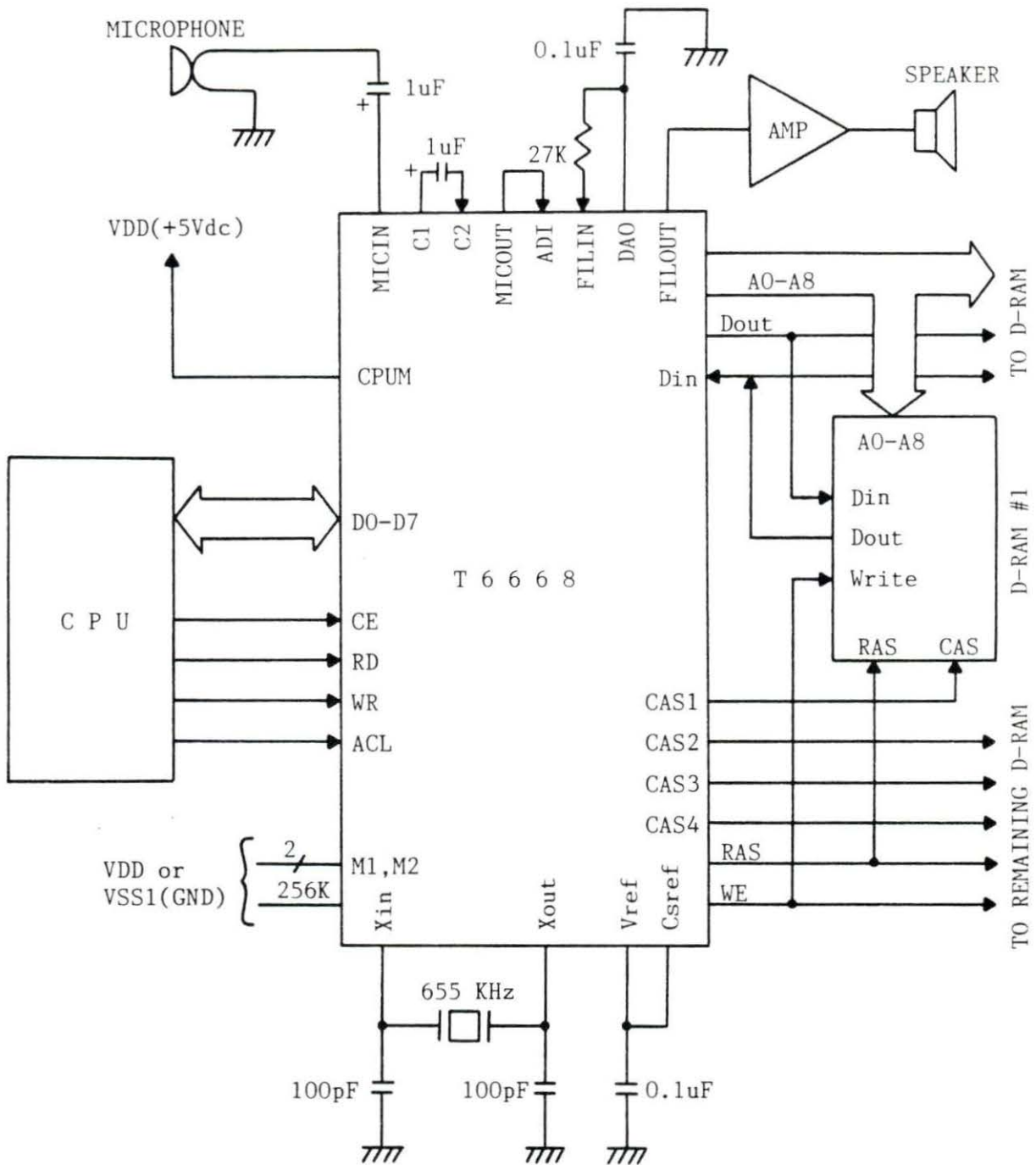
Figure 11. The T6668 used in the CPU mode (Toshiba America, Inc., 1988)

when an undefined command is given to the T6668. The NOP
command resets the ERR bit. The M2 and M1 bits have the same
values as pins M2 and M1 of the T6668, and their values are
based on the number of D-RAMS used. The last two status
register bits corresponding to pins D1 and D0 do not indicate
status and are always set low.

Table 3. Designation of status register bits (Toshiba
America, Inc., 1988)

D7	D6	D5	D4	D3	D2	D1	D0
BUSY	EOS	ERR	OVR	M2	M1	0	0

Audio circuit

The output of the bandpass filter (FILOUT) is connected
to the input of an audio amplification circuit before being
sent to the speaker during voice reproduction. The audio
circuit and its connections to the T6668 are shown in Figure
12.

The LM386 is a low-voltage audio amplifier that provides
a gain of 200 when pins 1 and 8 are connected by a capacitor.
The positive input (pin 3) is connected to a 20K potentiometer
to control the volume. The DVM-1 does not include a speaker,
but does provide a 3.5 mm jack for an external speaker
connection. The locations of the volume control and speaker
jack are shown in Figure 1.

Power circuit

Power to the DVM-1 is applied to a jack which accepts a 5.5mm X 2.0mm plug. The power requirement is 9-18 volts DC for the audio circuit; however, a transistor circuit is used to reduce this voltage to 5 volts DC for powering the T6668 and D-RAM. The location of the power jack is shown in Figure 1.
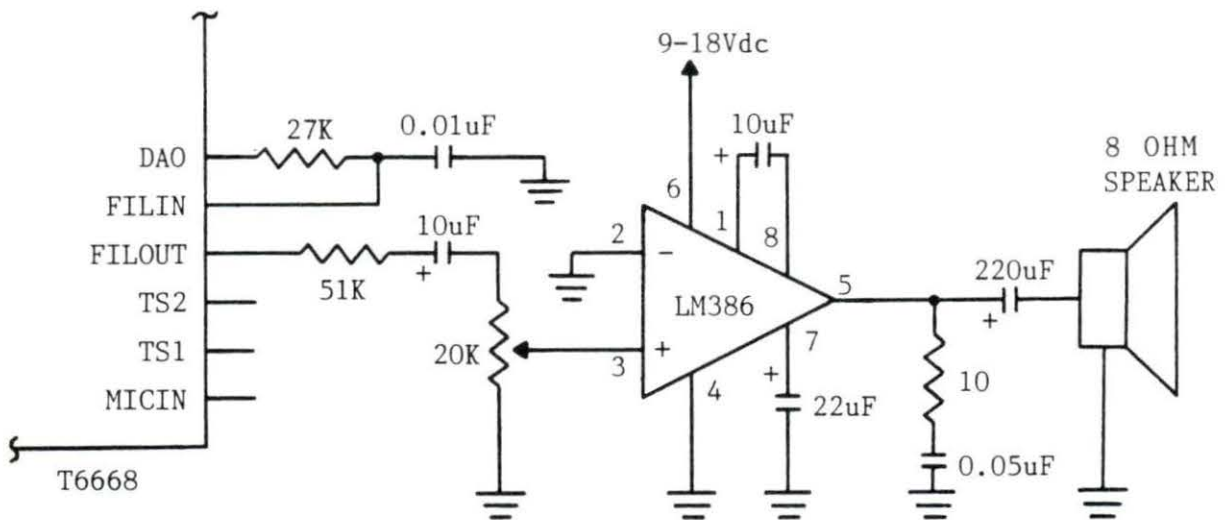
Figure 12. The audio circuit and its connection to the T6668

Memory

The memory of the DVM-1 consists of one to four pieces of either 64 Kbit or 256 Kbit dynamic RAM (D-RAM). Each bit of information is stored in the D-RAM using an FET capacitor, and due to leakage, the capacitors must be recharged or refreshed to retain memory (Lesea and Zaks, 1979). The pin diagram and nomenclature for a 256 Kbit D-RAM are shown in Figure 13.

Pins A0 to A8 address 262,144 memory locations. Pins A0 to A7 specify 256 rows or columns, and pin A8 specifies one of two sets of 256 rows or columns (Texas Instruments, Inc., 1984). The row-address strobe (RAS) latches bits A0 to A8 as the row address, and the column-address strobe (CAS) latches A0 to A8 as the column address.

				PIN NOMENCLATURE	
A8	1	16	VSS	A0–A8	Address Inputs
D	2	15	CAS	CAS	Column Address Strobe
W	3	14	Q	D	Data-in
RAS	4	13	A6	Q	Data-out
A0	5	12	A3	RAS	Row Address Strobe
A2	6	11	A4	W	Write Enable
A1	7	10	A5	VDD	+5-V Supply
VDD	8	9	A7	VSS	Ground
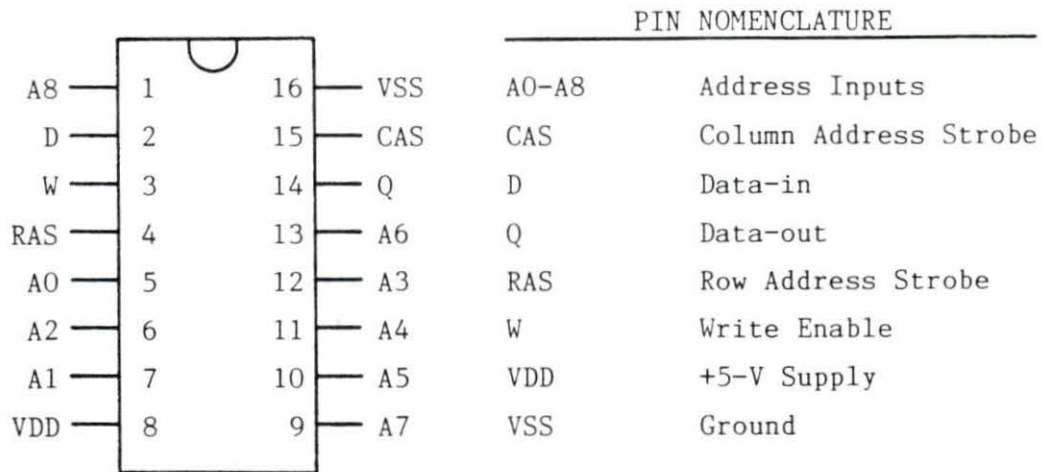
Figure 13. Pin diagram and nomenclature for a 256 Kbit D-RAM (Texas Instruments, Inc., 1984)

The write enable (W) input selects the read or write mode. The read mode is selected when the write enable input is set high, and the write mode is selected when the input is low.

The data-in pin (D) receives data during a write operation, and data exit the data-out pin (Q) during a read operation.

Pin V_{DD} is connected to a +5 volt DC supply, and V_{SS} is the ground pin.

The 256 Kbit D-RAM must be refreshed at least once every four milliseconds by strobing each of the 256 rows specified by bits A0 to A7 (Texas Instruments, Inc., 1984).

Control switches

Since the DVM-1 uses the T6668 in manual mode, switches are needed to control the device. The switches are shown in Figure 1; they include an 8-position SPST DIP switch and three momentary contact push-button switches.

DIP switch   Switches 1 and 2 (labeled Bit0 and Bit1, respectively) are used to select the recording bit rate according to Table 4. Switches 3, 4, 5, and 6 (labeled PH3, PH2, PH1, and PH0, respectively) are used to designate one of sixteen independent phrase recordings, as shown in Table 5. Switches 7 and 8 (labeled M1 and M2, respectively) indicate the number of D-RAMs installed, as indicated in Table 6.

Table 4.   Selection of recording bit rate using switches 1 and 2

Recording Bit Rate	Switch 1	Switch 2
8 Kbits/sec	OFF	OFF
11 Kbits/sec	ON	ON
16 Kbits/sec	OFF	ON
32 Kbits/sec	ON	ON

Push-button switches   The push-button switches are labeled RESET, RECORD, and PLAY. The reset switch is pressed

Table 5. Designation of phrase number using switches 3, 4, 5, and 6

Phrase Number	Switch 3	Switch 4	Switch 5	Switch 6
0	OFF	OFF	OFF	OFF
1	OFF	OFF	OFF	ON
2	OFF	OFF	ON	OFF
3	OFF	OFF	ON	ON
4	OFF	ON	OFF	OFF
5	OFF	ON	OFF	ON
6	OFF	ON	ON	OFF
7	OFF	ON	ON	ON
8	ON	OFF	OFF	OFF
9	ON	OFF	OFF	ON
10	ON	OFF	ON	OFF
11	ON	OFF	ON	ON
12	ON	ON	OFF	OFF
13	ON	ON	OFF	ON
14	ON	ON	ON	OFF
15	ON	ON	ON	ON

Table 6. Indication of the number of D-RAMs installed based on switch 7 and 8 settings

Number of D-RAMs Installed	Switch 7	Switch 8
1	OFF	OFF
2	OFF	ON
3	ON	OFF
4	ON	ON

to reset the address counter allowing a new recording to be made in place of a previous recording. Depressing the record switch starts the recording process, and releasing the switch stops recording. Momentarily depressing the play switch causes the selected phrase to be reproduced at the chosen bit rate.

The MC-1Z Single-Board Computer

The MC-1Z by Basicon, Inc. is a single-board computer intended for real-time process control applications. The MC-1Z is inexpensive, compact, and easy to use. The main components of the MC-1Z include:

1. The Zilog Z8671 microcomputer

2. The Intel 8255A programmable peripheral interface (PPI)

3. The National Semiconductor MM58274 programmable real-time clock/calendar

4. A 2K by 8 or an 8K by 8 CMOS RAM

5. An application socket for an EPROM or RAM expansion

6. The RS232 serial interface buffer

7. A power supply module

8. A negative voltage generator

A block diagram of the MC-1Z is given in Figure 14, and a memory map is shown in Figure 15.

The Zilog Z8671 microcomputer

The Z8671 is an eight-bit microcomputer preprogrammed with a BASIC/DEBUG interpreter capable of accessing the internal registers and external memory (Zilog, Inc., 1984). In addition to 2K bytes of on-chip ROM for storing the BASIC/DEBUG interpreter, the Z8671 also has a 144-byte register file, an on-board UART, two counter/timers, and 32 I/O lines provided by ports 0, 1, 2, and 3. A block diagram of the Z8671 is given in Figure 16.
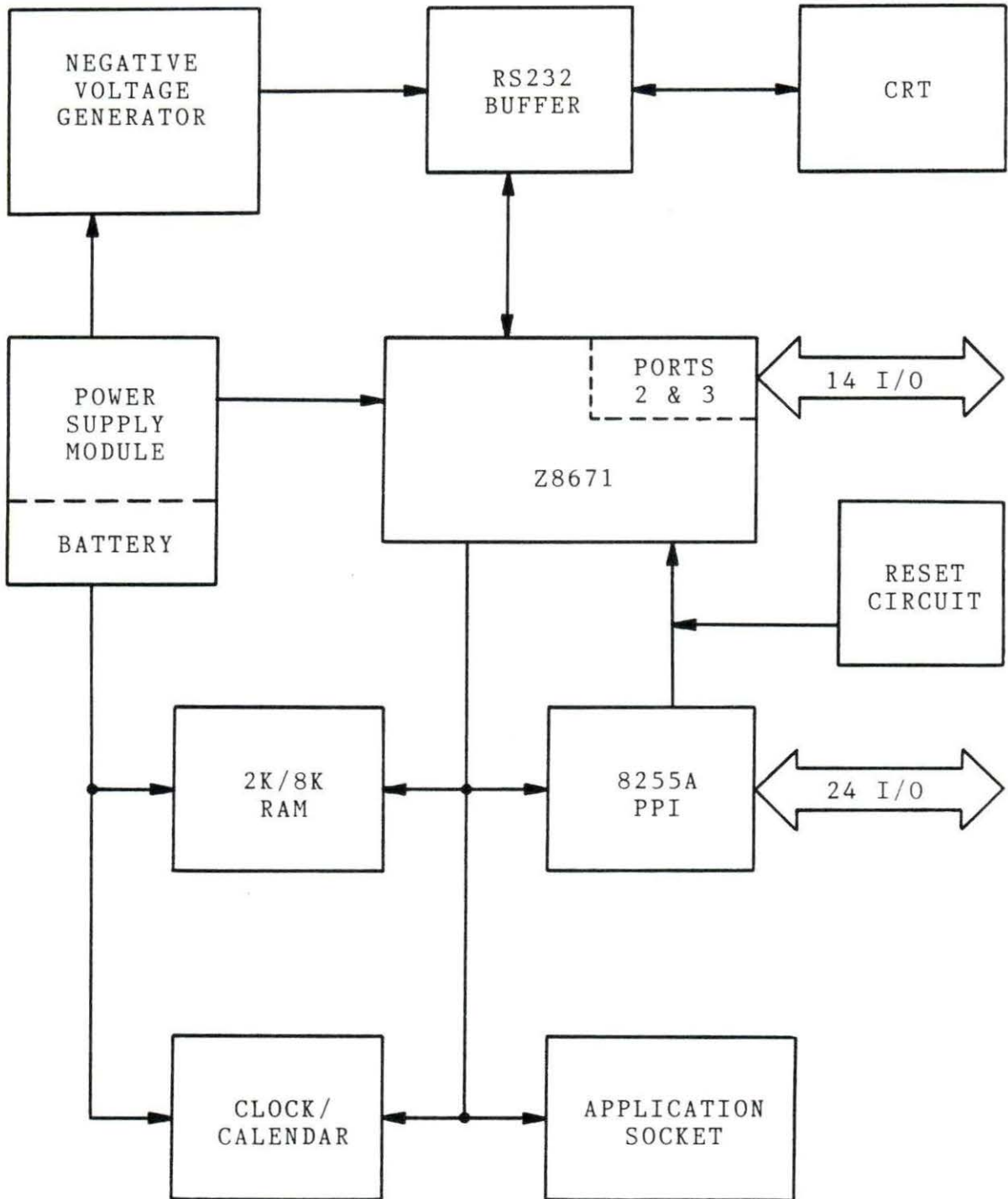
Figure 14.   Block diagram of the MC-1Z

HEXADECIMAL
ADDRESS

SIZE
(BYTES)

FFFD	BAUD RATE SELECT	1

B803		
	8255A PPI	4
B800		

A80F		
	CLOCK/CALENDAR	16
A800		

```
4FFF    Z8 VARIABLES (1/4K)
        - - - - - - - - - - - - - - - - -



                RAM                      8K



3000
```

```
1FFF
        EPROM (EXPANDABLE TO 8K)
                                         4K
      ─ AUTO-START AT 1020
1000
```

```
07FF
        Z8 BASIC AND REGISTERS           2K
0000
```

Figure 15. Memory map of the MC-1Z
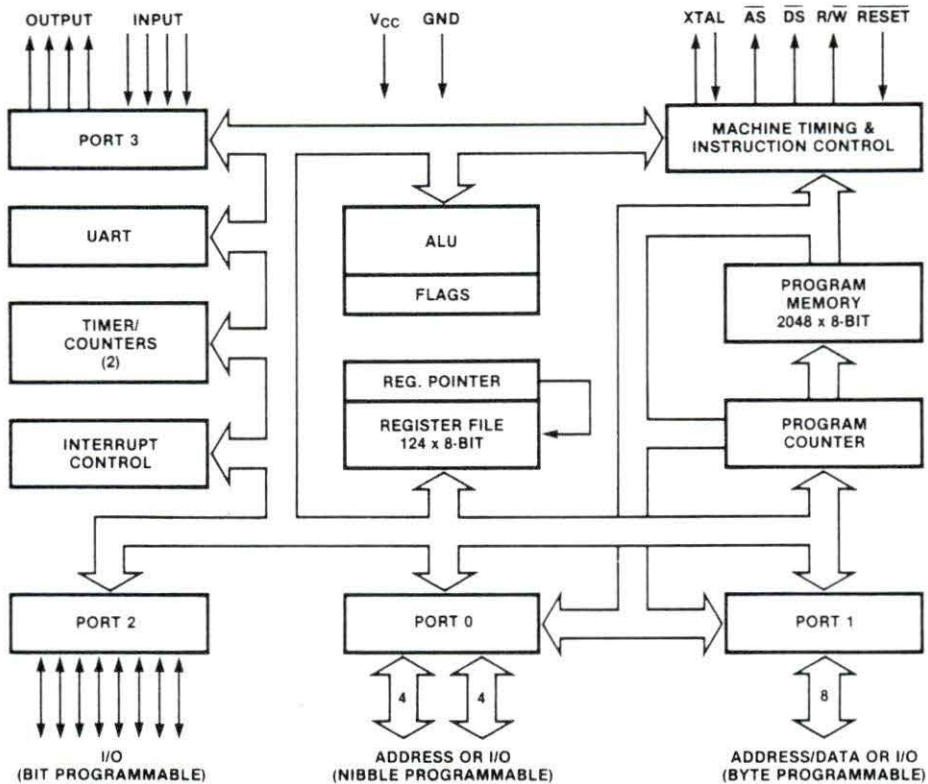
Figure 16. Block diagram of the Z8671 (Zilog, Inc., 1984)

BASIC/DEBUG interpreter   The BASIC/DEBUG interpreter recognizes a form of Dartmouth BASIC called BASIC/DEBUG. Since BASIC/DEBUG is intended for process control applications, Dartmouth BASIC capabilities such as trigonometric functions, arrays, and fractional numbers have been excluded (Zilog, Inc., 1981). Redundant commands and commands that can be accomplished using combinations of other statements have also been eliminated to conserve memory space. BASIC/DEBUG supports 26 variables, 10 operators, 2 functions, and 15 commands.

Variables   The 26 available variables are represented by each letter of the alphabet. Each variable occupies two bytes of RAM for storing numerical values (Zilog, Inc., 1981).

Operators   Operators signify a calculation that may be arithmetic or relational. The arithmetic operators include addition (+), subtraction (-), multiplication (*), and division (/). The relational operators are listed in Table 7.

Table 7. Relational operators supported by BASIC/DEBUG (Zilog, Inc., 1981)

SYMBOL	MEANING
=	equal
<=	less than or equal
<	less than
<>	not equal
>	greater than
>=	greater than or equal

Functions   The two functions are AND, for performing logical AND, and USR, for accessing machine language subroutines (Zilog, Inc., 1981).

Commands   The commands are listed and briefly explained in Table 8.

Ports 2 and 3   Port 2 provides 8 I/O lines that can be independently set for input or output using the Port 2 mode register located at address 246 (Basicon, Inc., 1984). As

Table 8. Commands recognized by BASIC/DEBUG

COMMAND	FUNCTION
GO@	Unconditional branching to a machine language subroutine
GOSUB	Unconditional branching to a BASIC subroutine
GOTO	Unconditional branching within a program
IF/THEN	Conditional branching and operations
INPUT, IN	Data entry
LET	Value assignment to a variable or memory location
LIST	Displays memory on the CRT screen
NEW	Clears memory for a new program
PRINT, PRINTHEX	Displays characters and/or numerical values on the CRT screen
REM	Signifies a program remark
RETURN	Indicates the end of a BASIC subroutine
RUN	Starts program execution
STOP	Ends program execution

shown in Figure 17, a Port 2 mode register bit set high designates the corresponding Port 2 bit as an input, and a bit set low produces an output line. By equating address 246 (@246) to a decimal value between 0 and 255 or a hexadecimal value between %00 and %FF, Port 2 can be configured in any possible I/O combination. For example:

```
@246=%00 or 0        D7-D0 are output lines

@246=%0F or 15       D7-D4 are output lines
                     D3-D0 are input lines

@246=%F0 or 240      D7-D4 are input lines
                     D3-D0 are output lines

@246=%FF or 255      D7-D0 are input lines
```

D7	D6	D5	D4	D3	D2	D1	D0
X	X	X	X	X	X	X	X
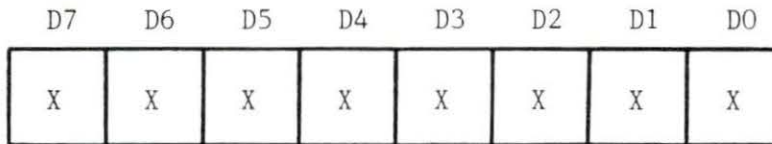
Note: X=0, output. X=1, input.

Figure 17.   The Port 2 mode register (Zilog, Inc., 1984)

Port 3 provides 4 input lines (D0-D3) and 4 output lines
(D4-D7).   In addition to the I/O function, Port 3 lines can be
configured for handshake I/O, interrupt requests, serial I/O,
and counting/timing.   The various functions are selected by
writing a control word to the Port 3 mode register located at
address 247 (@247).   For example, if @247=%41 then Port 3 is
set for input on lines D0-D3 and output on lines D4-D7.   In
the MC-1Z system, lines D0 and D7 are used for serial
communication with a terminal leaving 6 lines (D1-D6) of Port
3 directly available to the user.

The 8255A programmable peripheral interface (PPI)

The Intel 8255A PPI provides 24 input/output lines that
supply ports A, B, and C, each having 8 lines or bits.   The

PPI also has a control register to program the ports for input or output. The high and low nibbles of port C can be programmed independently.

In the MC-1Z system, the 8255A PPI is located in four successive memory addresses:

%B800	Port A
%B801	Port B
%B802	Port C
%B803	PPI control register

Ports A, B, and C are set for input or output by writing to the control register at location %B803. The control register is shown in Figure 18, and the possible input/output combinations for the ports are listed in Table 9.

D7	D6	D5	D4	D3	D2	D1	D0
1	0	0	X	X	0	X	X

Note: X=0, output. X=1, input.

Figure 18.   The control register for setting ports A, B, and C for input and/or output (Uffenbeck, 1985)

If the 8255A PPI is reset, the ports are set for input and the lines are floating. When a word is written to the control register, all chosen output lines are set to logic 0 (Basicon, Inc., 1984).

Table 9. Possible I/O combinations for ports A, B, and C
(Basicon, Inc., 1984)

Hexadecimal Value of @%B803	Port A	Port B	Port C High Nibble	Port C Low Nibble
%80	OUTPUT	OUTPUT	OUTPUT	OUTPUT
%81	OUTPUT	OUTPUT	OUTPUT	INPUT
%82	OUTPUT	INPUT	OUTPUT	OUTPUT
%83	OUTPUT	INPUT	OUTPUT	INPUT
%88	OUTPUT	OUTPUT	INPUT	OUTPUT
%89	OUTPUT	OUTPUT	INPUT	INPUT
%8A	OUTPUT	INPUT	INPUT	OUTPUT
%8B	OUTPUT	INPUT	INPUT	INPUT
%90	INPUT	OUTPUT	OUTPUT	OUTPUT
%91	INPUT	OUTPUT	OUTPUT	INPUT
%92	INPUT	INPUT	OUTPUT	OUTPUT
%93	INPUT	INPUT	OUTPUT	INPUT
%98	INPUT	OUTPUT	INPUT	OUTPUT
%99	INPUT	OUTPUT	INPUT	INPUT
%9A	INPUT	INPUT	INPUT	OUTPUT
%9B	INPUT	INPUT	INPUT	INPUT

The MM58274 clock/calendar

The National Semiconductor MM58274 provides a program-mable, real-time clock/calendar for the MC-1Z system. The clock measures tenths of seconds through years and also accounts for leap years (Basicon, Inc., 1984). The clock/calendar counters are four-bit counters that use bits D0-D3 of the data bus. The various counters and their respective memory addresses and modes are listed in Table 10.

2K/8K RAM socket

The MC-1Z has a socket for accepting a 2K by 8 or an 8K by 8 CMOS RAM. With memory addresses %3000 through %4FFF

reserved for RAM, the MC-1Z is ready to operate with an 8K RAM provided jumper E4 is installed (Basicon, Inc., 1984). To use a 2K RAM, jumper E5 must be installed, and a short sequence of data must be written to the Z8 registers to allow the MC-1Z to adjust to the change in memory (Basicon, Inc., 1984). The MC-1Z system used in this project has 8K bytes of RAM for storing programs; however, the top 1/4K bytes are designated for BASIC/DEBUG variables and other functions (Basicon, Inc., 1984).

Table 10. Clock counters and their addresses and modes (Basicon, Inc., 1984)

Address	Counter	Mode
%A800	Control register	Split read and write
%A801	Tenths of seconds	Read only
%A802	Units of seconds	Read or write
%A803	Tens of seconds	Read or write
%A804	Units of minutes	Read or write
%A805	Tens of minutes	Read or write
%A806	Units of hours	Read or write
%A807	Tens of hours	Read or write
%A808	Units of hours	Read or write
%A809	Tens of hours	Read or write
%A80A	Units of months	Read or write
%A80B	Tens of months	Read or write
%A80C	Units of years	Read or write
%A80D	Tens of years	Read or write
%A80E	Day of week	Read or write
%A80F	Clock setting/ interrupt reg.	Read or write

The MC-1Z application socket

The application socket accepts a 4K/8K EPROM or a 2K/8K RAM (Basicon, Inc., 1984). The memory allocated for the application socket exists between addresses %1000 and %2FFF, but since this MC-1Z system uses a 4K EPROM, only memory between %1000 and %1FFF is used. One feature of the Z8671 allows a program to automatically start when the system is powered. For this feature to work, the program must start at address %1020 and contain a line number between 0 and 255 followed by a BASIC program statement (Basicon, Inc., 1984).

The RS-232C serial interface buffer

To standardize connections between terminals and modems, the RS-232C standard was developed in the early 1960s to define logic levels, maximum baud rates (bits/second), maximum cable lengths, and connector types (Uffenbeck, 1985). The RS-232C standard has a maximum baud rate of 20K for a 50 foot cable. The RS-232C standard defines a voltage between +3 and +25 as logic 0 and a voltage between -3 and -25 as logic 1. These logic levels provide a minimum of 2 V of noise immunity which is necessary for dealing with the capacitive and DC loading effects associated with long cables (Uffenbeck, 1985).

The MC-1Z has two RS-232C serial interface buffers, one for receiving and the other for transmitting. The receiver converts the incoming RS-232C logic signals to TTL-compatible signals that are then applied to the serial input line (D0) of

port 3 of the Z8671 microcomputer. The transmitter converts the serial output signal from port 3 (D7) into an RS-232C compatible signal that is sent to the CRT terminal. A ground wire completes the 3-wire serial communication system of the MC-1Z.

Power supply module

The MC-1Z requires a +5 Vdc (+/- 5%) power source that can deliver 250 mA or 500 mA when the EPROM programmer is used. The power supply module provides +5 Vdc to the Z8671 microcomputer and the negative voltage generator discussed in the next section. The power supply module also has a 100 mA-hr., 3.6 volt nickel-cadmium battery to maintain programs stored in RAM and to keep the calendar/clock running when power is interrupted (Basicon, Inc., 1984). The battery is trickle-charged through a charging resistor included with the MC-1Z. The power supply module also contains a DB25S connector allowing connection to any standard RS-232 cable, a push-button switch for resetting the system without disturbing the RAM contents, a power ON-OFF switch, and an LED power indicator.

Negative voltage generator

Powered by the power supply module, the negative voltage generator converts +5 Vdc to -5 Vdc for use by the RS-232 serial interface buffer.

Interfacing the DVM-1 to the MC-1Z

Interfacing the digital voice recorder (DVM-1) to the single-board computer (MC-1Z) involves alterations to the DVM-1 board and development of an interface circuit.

DVM-1 alterations

Alterations to the DVM-1 are necessary to make the control pins of the T6668 (RD, WR, CE, ACL, CPUM, AND D0-D7) available for direct connection to the ribbon cable from the interface circuit. Pins D0, D1, D2, D3, D6, and D7 are connected to the off side of the DIP switch. Wires soldered to this side can be used to access the pins provided the switches are off to insure electrical isolation from the original circuit.

Pins D4, D5, WR, and ACL are also connected to components on the DVM-1 board. These pins are electrically isolated by breaking the path between the pin and the components to which they were connected. Wires are then soldered on each side of the break allowing the circuit to be reconnected in its original form.

Since the CPUM pin is soldered directly to ground for manual control, it is disconnected by heating the soldered connection until the pin can be pried from the DVM-1 board. To avoid overheating the T6668 and applying too much solder, a conductive adhesive is used to glue a wire to the pin. The glueing process involves the following steps:

1. Prepare the surfaces to be glued by cleaning them with ethanol.

2. Mix equal parts of the two-component conductive adhesive.

3. Dip the stripped end of a 30-gauge wire into the conductive adhesive mixture to coat the wire.

4. Lay this end of the wire onto the pin of the T6668.

5. Hold the wire in place with a wirewrap post clamped into a hemostat. The notched end of the post sets on the wire, and the weight of the hemostat holds the wire down until the adhesive sets overnight.

6. The remaining length of the wire is secured to the DVM-1 board using drops of 5-minute epoxy.

Pin CE is handled in the same manner as pin CPUM. Since pin RD is not used in the manual mode, it is accessed by simply glueing a wire to the pin. All of the wires, as well as wires from +5 volts and ground, are soldered to a 2 X 13 pin header mounted to the DVM-1 board. The circuit alterations are shown schematically in Figure 19.

The interface circuit

The interface circuit consists of a ribbon cable for connection to the DVM-1 pin header, control switches for selecting manual or CPU control, and a 2 X 13 pin header for connection to the ribbon cable from the PPI register.

The control switches are SPDT toggle switches that allow the DVM-1 board to be operated in the manual mode as originally intended or by CPU control. The switches are associated with pins D4, D5, WR, and ACL which are disconnected from components on the DVM-1 board by breaking

the electrical path between the T6668 and the components.
Wires soldered to the T6668 side of the break are labeled D4,
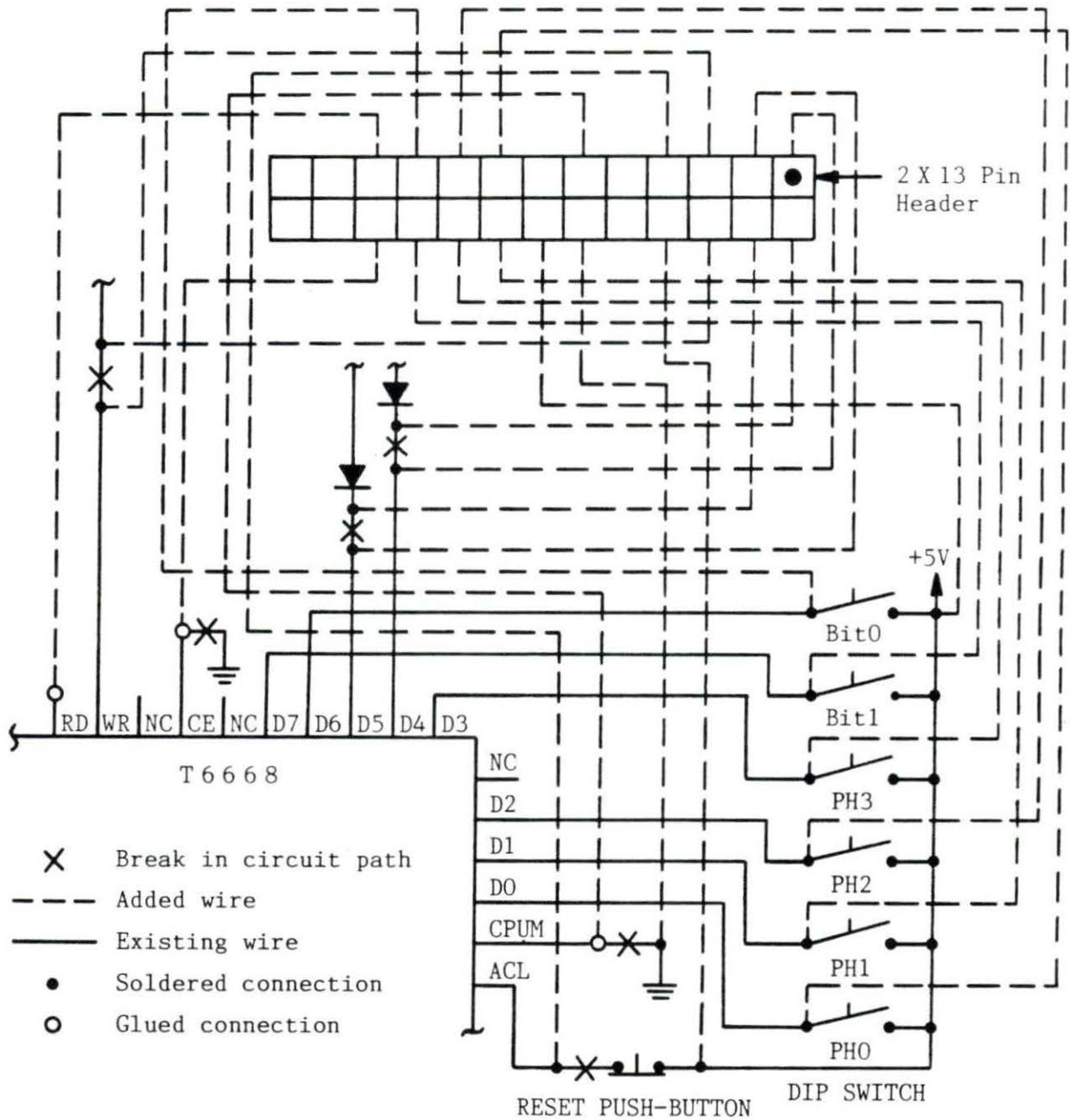
Figure 19. Circuit alterations of the DVM-1 board

D5, WR, and ACL and are connected to the center post of the SPDT switch. The corresponding wires soldered on the other side of the break are labeled D4', D5', WR', and ACL' and are each connected to one of two remaining posts of the toggle switch. This arrangement allows the break in the electrical path to be bypassed returning the circuit to its original form.

The remaining post of each toggle switch is wired to a 2 X 13 pin header for connection to the PPI register. When the switches connect pins D4, D5, WR, and ACL to the pin header, computer control is possible. Pin CPUM is also wired to a toggle switch with ground wired to one post and +5 Vdc wired to the other post. As shown in Figure 20, pins D0-D7 are connected to port A (A0-A7), and pins ACL, WR, RD, and CE are connected to bits B3, B2, B1, and B0, respectively.

Software

The software allows the MC-1Z to control the DVM-1 by writing commands to the T6668 microprocessor. A listing of the T6668 commands is given in Table 11. All of the commands require 8 bits (1 byte) except for the ADLD1 and ADLD2 commands which require 24 bits (3 bytes) for addressing all of the bits in memory.

The commands are written to the T6668 through port A of the programmable peripheral interface (PPI). The logic level of control pins ACL, WR, RD, and CE are set by bits B3, B2,

B1, and B0 of port B, respectively. To signify that a command
is going to be written, pins CE and WR must be set low. The
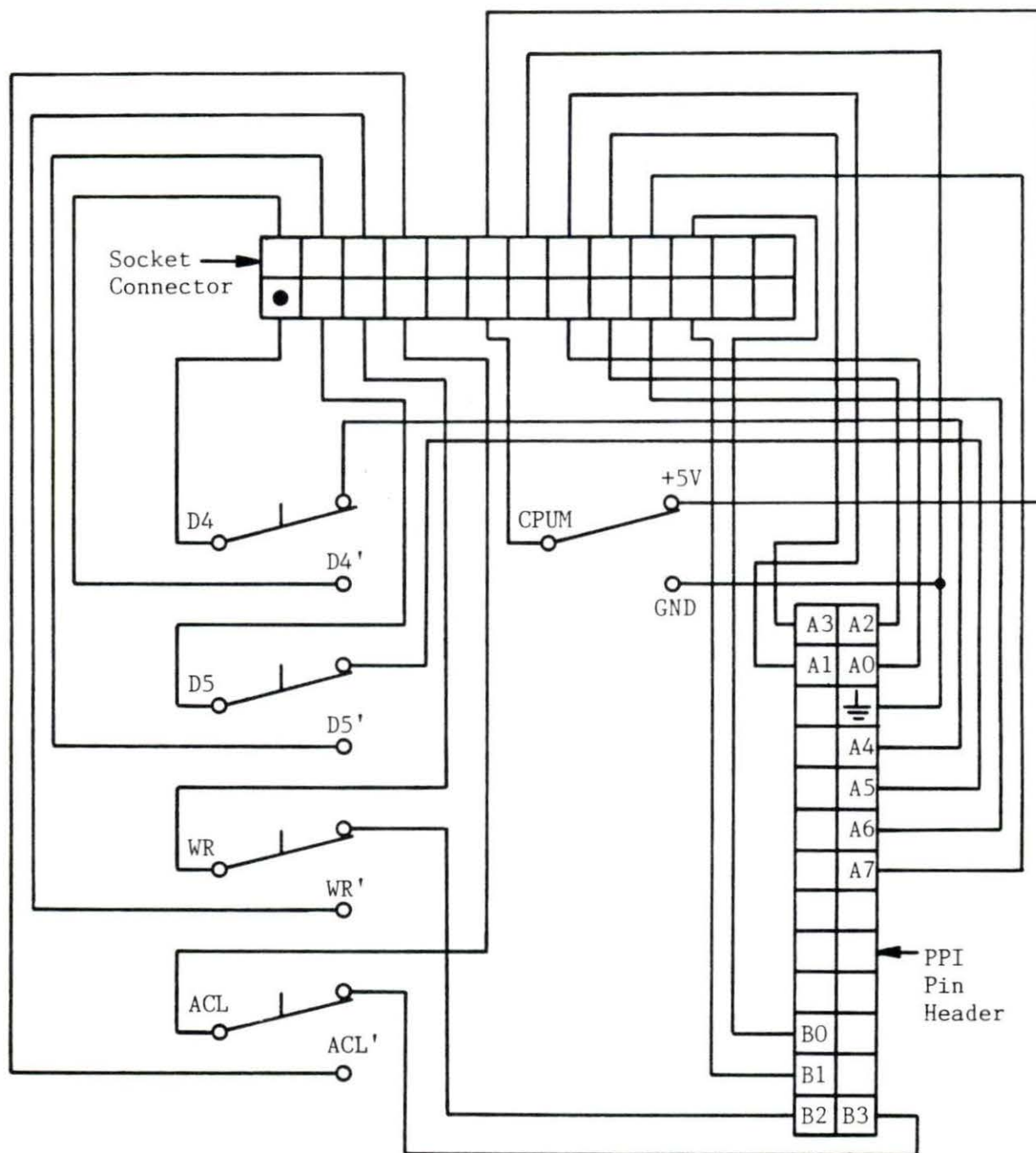
Figure 20. The interface circuit

Table 11. T6668 commands (Toshiba America, Inc., 1988)

Command	D7	D6	D5	D4	D3	D2	D1	D0	Function
NOP	0	0	0	0	X	X	X	X	Selects the sound reproduction mode
START	0	0	0	1	X	X	X	X	Starts recording/re-production in the Direct mode
STOP	0	0	1	0	X	X	X	X	Stops sound recording/reproduction in the Label/Index mode
ADLD1	0	0	1	1	A19	A18	A17	A16	Specifies the start address in the Direct mode(3 bytes)
	A15	A14	A13	A12	A11	A10	A9	A8	
	A7	A6	A5	A4	A3	A2	A1	A0	
ADLD2	0	1	0	0	A19	A18	A17	A16	Specifies the end address in the Direct mode(3 bytes)
	A15	A14	A13	A12	A11	A10	A9	A8	
	A7	A6	A5	A4	A3	A2	A1	A0	
CNDT	0	1	0	1	X	SL	BR1	BR0	SL selects sounds(0) or silent(1) mode. BR1 and BR0 specify bit rate(bits/sec.)
LABEL	0	1	1	0	LB3	LB2	LB1	LB0	Starts recording/reproduction in the Label/Index mode
ADRD	0	1	1	1	X	X	X	X	Allows the address counter to be read
	0	0	0	0	A19	A18	A17	A16	in 3 successive read
	A15	A14	A13	A12	A11	A10	A9	A8	operations as shown
	A7	A6	A5	A4	A3	A2	A1	A0	to the left
REC	1	0	0	0	X	X	X	X	Selects the sound recording mode

X = don't care.

command is actually written after pins CE and WR are returned
to their normal state of logic 1 (Toshiba America, Inc.,
1988). The commands provide two different recording/
reproduction modes called the Label/Index mode and Direct
mode.

Label/Index mode

As with manual control, the Label/Index mode uses part of
the memory as an index area for storing start and stop
addresses and is also capable of storing 16 different
recordings. Recording is accomplished as follows:

1. Resetting the address counter of the T6668 by setting
   the ACL pin low and then high.

2. Inputting the REC command to set the T6668 in the
   recording mode.

3. Entering the CNDT command to specify the bit rate.

4. Inputting the LABEL command to indicate phrase
   number.

Recording begins after the LABEL command is entered and stops
when the STOP command is given. The corresponding port values
and command statements for each step of the recording process
are given in Table 12. A sample program is listed in Appendix
B.

Reproduction in the Label/Index mode is accomplished by:

1. Entering the NOP command to set the T6668 into the
   reproduction mode.

2. Inputting the LABEL command to specify which phrase
   is to be reproduced.

Reproduction begins after the LABEL command is entered and stops when the recording ends. The command sequence is:

    @A=%00: @B=10: @B=15 (writing the NOP command)

    @A=%68: @B=10: @B=15 (writing the LABEL command)

Table 12. Function, port value, and command statement for each step of the recording process in the Label/Index mode

Function or Command	Port A or B Values D7----------D0	Command Statement
	X X X X A W R C	
Reset the address counter	0 0 0 0 0 1 1 1	@B=7 or @B=%07
	0 0 0 0 1 1 1 1	@B=15 or @B=%0F
	7 6 5 4 3 2 1 0	
REC *	1 0 0 0 0 0 0 0	@A=128 or @A=%80
CNDT(32K) *	0 1 0 1 0 0 1 1	@A=83 or @A=%53
LABEL(#8) *	0 1 1 0 1 0 0 0	@A=104 or @A=%68
STOP *	0 0 1 0 0 0 0 0	@A=32 or @A=%20
	X X X X A W R C	
*Writes command to the T6668	0 0 0 0 1 0 1 0	@B=10 or @B=%0A
	0 0 0 0 1 1 1 1	@B=15 or @B=%0F

        X = don't care   A = ACL (Reset)   W = WR (Write)
        R = RD (Read)   C = CE (Chip Enable)

        % indicates a hexadecimal number
        @ refers to a memory location

Direct mode

In the Direct mode, the ADLD1 and ADLD2 commands specify the memory start and stop addresses, respectively. The Direct mode has two advantages compared to the Label/Index mode:

1. Since the start and stop addresses are specified by commands, an index area is not needed allowing more memory space for voice storage.

2. By designating start and stop addresses, an indefinite number of phrases can be independently recorded compared to the maximum of 16 phrases allowed in the Label/Index mode.

The recording process is started by following these steps:

1. Setting the ACL pin low and then high to reset the address counter

2. Inputting the REC command to set the T6668 into the recording mode.

3. Inputting the CNDT command to specify the bit rate.

4. Entering the ADLD1 and ADLD2 commands to designate the start and stop addresses, respectively.

5. Inputting the START command to begin voice recording.

Recording stops when the stop address is reached. The command sequence is similar to the previous example except that commands ADLD1 and ADLD2 each require three write sequences since they each occupy three bytes. The high nibble of the first byte indicates whether the address is a start or stop address. The remaining 2 and 1/2 bytes (20 bits) are used to address the memory. A sample program is given in Appendix C.

Reproduction in the Direct mode involves the following steps:

1. Inputting the NOP command to set the T6668 into the reproduction mode.

2. Entering the ADLD1 and ADLD2 commands to specify the region of memory to be reproduced.

3. Inputting the START command to begin voice reproduction.

Reproduction stops at the stop address.

Firmware

The firmware is a utility PROM (ZUTIL-1.00) that supports the MC-1Z by offering 14 commands for program development. The PROM consists of a main program written in BASIC and several machine code subroutines for fast execution of commands. With the PROM installed in the MC-1Z application socket, the utility program automatically starts running when the MC-1Z system is powered or reset. The utility program is stopped by pressing the ESCape key and then the RETURN key. The 14 available commands and their functions are listed in Table 13.

Two situations for using these commands include preparing the RAM before entering a BASIC program and transferring a program in RAM to an EPROM. To prepare the RAM, the F command fills the RAM with %FF, and the M command marks the beginning of the program. To transfer the program to an EPROM, the D command is first used to determine the RAM address where the program ends. This address is needed when specifying what part of the RAM gets copied to the EPROM. The E command checks the EPROM to see if it is fully erased. The P command

is then entered to start the copying process. The V command verifies whether or not the copying was successful. The R command displays the contents of the EPROM to see what was copied. Most of these utility commands include an argument consisting of a starting, ending, and/or destination address.

Table 13. Commands offered by the MC-1Z utility PROM

Command	Function
A (Alter memory)	Allows direct access to memory for changing bytes
C (Copy memory)	Copies any section of memory to another area of memory
D (Display memory)	Displays the RAM contents
F (Fill memory)	Fills any memory location with a chosen hexadecimal value
E (Erase check)	Checks EPROM bytes to see if they equal %FF
P (Program EPROM)	Copies RAM contents into EPROM
V (Verify EPROM)	Verifies whether or not EPROM contents match the RAM contents
R (ROM display)	Displays contents of EPROM
M (Mark top)	Marks the beginning of a program
L (Locate)	Searches memory for a selected hexadecimal number
H (Help)	Displays a list of utility commands
S (Set time)	Sets the clock/calendar
T (Time check)	Displays the time and date

SPEECH RECOGNITION

Literature Review

Speech recognition is the process of differentiating words in a vocabulary spoken by the same person or different people. The following discussion of speech recognition includes the problems of accomplishing speech recognition, the categories of speech recognition, the speech recognition process, and applications of speech recognition for the disabled.

Problems

The problems of speech recognition include ambiguities of speech and variations in pronunciation between different speakers or from the same speaker.

Ambiguities of speech   One of the ambiguities of speech results from the pronunciation of words in continuous speech. Continuous speech is the pronunciation of words without pauses between the words in which the individual words can still be discerned by the individual (Wallich, 1987). The ambiguity arises when the pronunciation of a group of adjacent words suggests more than one possible combination of words. For example, a speech recognizer would have difficulty distinguishing between "grey tape" and "great ape," or between "sixteen ages" and "six teenagers" (Hollingum and Cassford, 1988).

One way to avoid the ambiguities of continuous speech is to "isolate" words during speech. Isolated words are words spoken with a long enough pause between them so that the pronunciation of a word is not affected by words immediately before or after it (Wallich, 1987). Unfortunately, this style of speech also has ambiguities due to words with the same pronunciation called homonyms. Examples of homonyms include "great" and "grate," and "to," "two," and "too."

<u>Variations in pronunciation</u>   Dialects are one cause of variations in pronunciation between different speakers. For example, one person might pronounce the word "creek" as "creek" while another person would pronounce it as "crick." Accents are another source of variations in pronunciation.

Variations in pronunciation also occur with the same person. A person may change word pronunciation as a result of speaking faster. For example, "bread and butter" can be spoken faster by saying "bread 'n' butter," or even faster by saying "brembutter" (Hollingum and Cassford, 1988). Other factors affecting pronunciation include emotional status or a physical condition such as a cold.

Categories

The categories of speech recognition are based on various levels of accomplishment in overcoming the problems of speech recognition. The categories are listed below in order of increasing sophistication:

1. Isolated-word/speaker-dependent

2. Isolated-word/speaker-independent

3. Continuous-speech/speaker-dependent

4. Continuous-speech/speaker-independent

The first category is the least sophisticated because it only considers isolated words and speech from one person (speaker-dependent). The second category is more difficult because it includes speech input from any person (speaker-independent). The next two categories have increased capability because they allow continuous speech; however, the fourth category is more sophisticated because it accepts speech from anyone. Each of these categories can be further classified based on the number of words that can be recognized.

The speech recognition process

The speech recognition process begins with the production of speech sound from the human vocal tract. The speech sound is collected by a microphone for conversion into an electrical signal. Various speech signal acquisition and analysis techniques extract the necessary information used by the speech recognition algorithms to differentiate words.

Speech sound production   The sounds of speech are controlled by the entire vocal tract shown in Figure 21. Speech sound begins when air exhaled from the lungs passes through the vocal chords causing them to vibrate. The vibration of the vocal chords is referred to as glottal

vibration which contributes to voice pitch. The speech sound is then altered as it passes through the throat (pharynx), oral, and nasal cavities. These cavities act as resonators to produce various tones much like the pipes of a pipe organ (Cater, 1984). In addition to the vocal chords and resonant
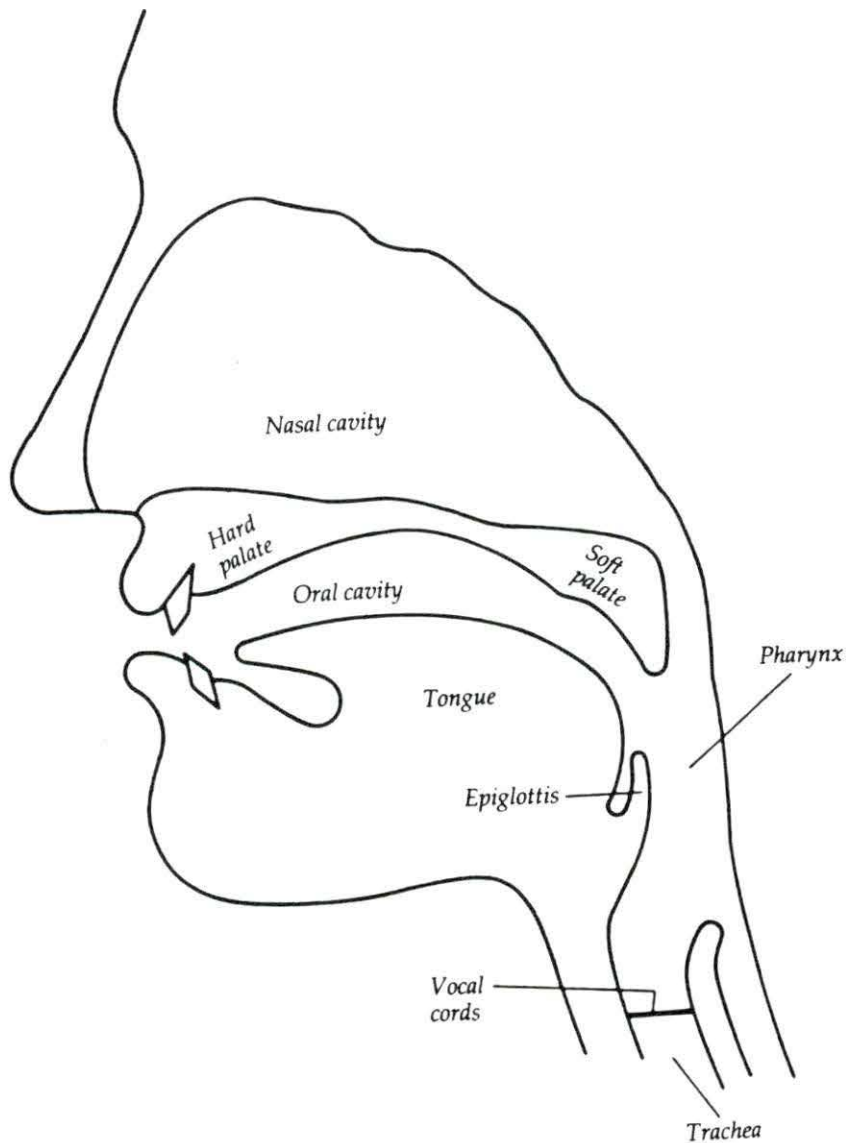
Figure 21. Anatomy of the human vocal tract (Hollingum and Cassford, 1988)

cavities, the tongue, palate, teeth, and lips also affect the speech sound to enable the pronunciation of all the basic sounds of a language. These basic sounds are called phonemes.

The phonemes are distinguishable by the resonances produced by the cavities. Each of the three cavities produces a unique resonant frequency for a particular sound. This frequency is called a formant frequency which is evident in the speech spectrum shown in Figure 22. The first, second, and third formants are associated with the throat, nasal, and oral cavities, respectively (Cater, 1984). Although other formants may be present, they do not contribute significantly to the total energy of the speech spectrum. Depending on the
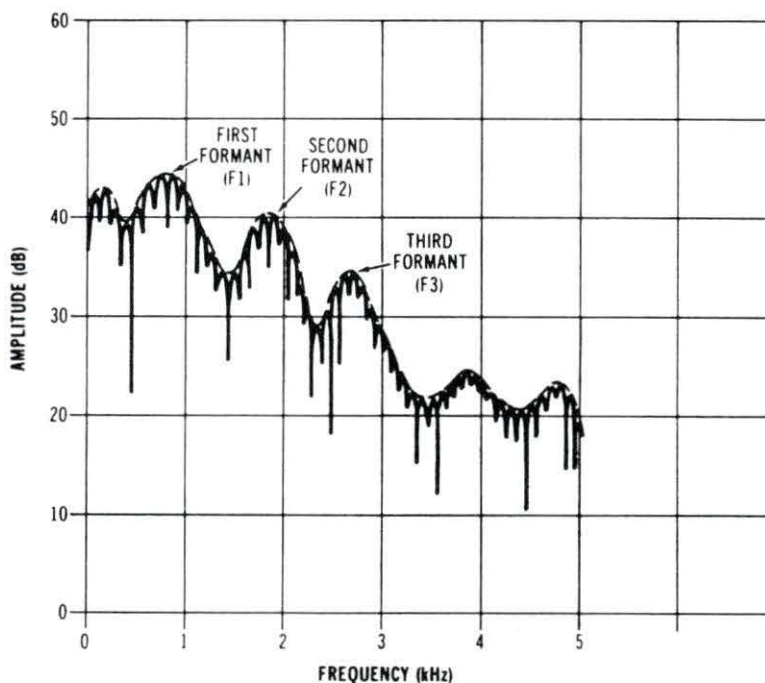
Figure 22. A speech spectrum showing the formants (Cater, 1984)

phoneme spoken, each formant will reside in a particular frequency band as indicated in Table 14 which lists the formant resonances of the vowel sounds.

Table 14. Formant resonances of vowel sounds (Cater, 1984)

Vowel	F1 Resonance(Hz)	F2 Resonance(Hz)	F3 Resonance(Hz)
ee(eat)	210-330	2230-2350	2950-3070
i(bit)	330-450	1930-2050	2490-2610
eh(bet)	470-590	1780-1900	2420-2540
ae(bat)	600-720	1660-1780	2350-2470
ah(top)	670-790	1030-1150	2380-2500
aw(ball)	510-630	780-900	2350-2470
oo(book)	380-500	960-1080	2180-2300
oo(moon)	240-360	810-930	2180-2300
uh(tug)	580-700	1130-1250	2330-2450
er(nerve)	430-550	1290-1410	1630-1750

Speech signal acquisition   The purpose of speech signal acquisition is to convert the acoustic speech signal, collected by the microphone, into a form that will allow features to be extracted.   These features enable sounds to be characterized for the purpose of speech recognition.   Two primary methods of speech signal acquisition are direct waveform acquisition and spectral signal acquisition (Cater, 1984).

Direct waveform acquisition   Direct waveform acquisition is the result of digitizing the analog speech signal.   One method of direct waveform acquisition is waveform encoding in which the analog signal is sampled and held until

an analog-to-digital converter can convert the signal into a voltage-equivalent binary data byte (Cater, 1984).

Delta modulation encoding is another method of directly digitizing the speech signal. Instead of representing each sample with its own data byte as in waveform encoding, delta modulation encoding only records the changes in the speech signal without concern for the actual voltage of the sample.

<u>Spectral signal acquisition</u>   Spectral signal acquisition obtains the speech frequencies directly from the speech signal by filtering the signal before performing any digitization. This method uses several bandpass filters that span the frequency range of the speech signal. The filters extract the various frequency components which are then digitized and stored in separate registers of the computer (Cater, 1984). The filtering can be analog, digital, or mathematical.

<u>Speech signal analysis</u>   Speech signal analysis extracts features from the information provided by speech signal acquisition. The main methods of feature extraction include the zero-crossing technique, Fast-Fourier Transform (FFT), and linear predictive coding (LPC).

<u>Zero-crossing technique</u>   This technique filters the speech signal to separate each formant frequency. A zero-crossing detector then counts the number of times that the output signal of each bandpass filter changes sign during a

set period of time. The counting is a measure of the extent
of each formant in a particular sound. The counter values are
the extracted features used in speech recognition.

Fast-Fourier Transform (FFT)   FFT is a
mathematical process that converts a time-varying signal
directly into a signal spectrum representing the energy of the
signal as a function of frequency. The process is illustrated
in Figure 23. A sequence of FFTs forms a spectrogram as shown
in Figure 24. The spectrogram acts as a "fingerprint" for
distinguishing sounds in speech recognition.
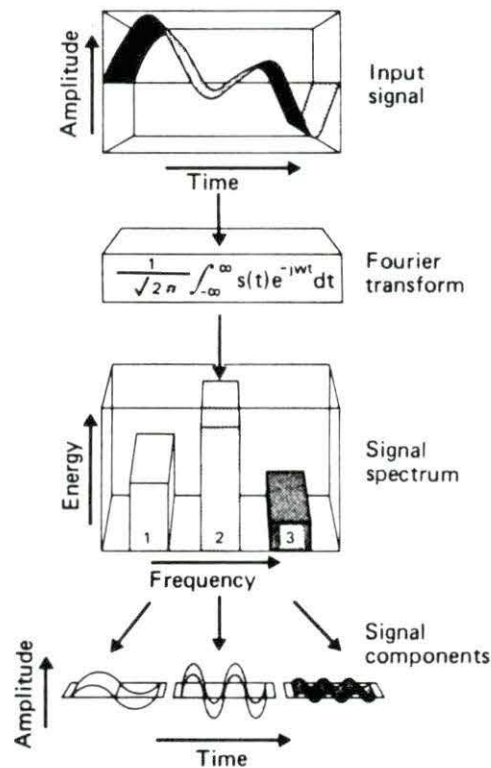
Figure 23.   Process of a Fast-Fourier Transform (Bristow,
1984)
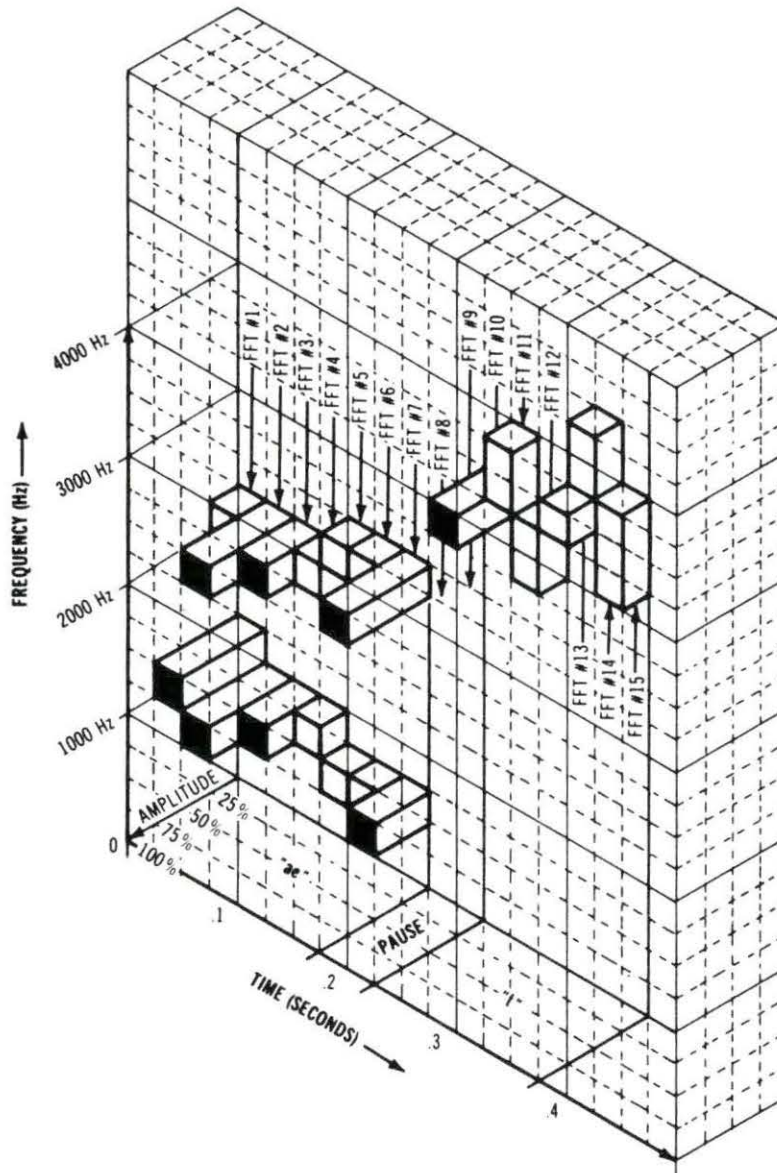
Figure 24. A spectrogram consisting of a series of FFTs
           (Cater, 1984)

Linear predictive coding (LPC)   Based on an

acoustical model of human speech, LPC predicts speech waveform

amplitude on the basis of a linear combination of amplitudes

from previous samples (Bristow, 1984). With LPC filtering of the speech signal, a series of LPC coefficients are produced. These coefficients can then be compared to distinguish words.

Speech recognition algorithms   Speech recognition algorithms compare extracted features in a process called template matching. Using extracted features, the algorithm develops a template for each word in the vocabulary and compares these templates to the template of the word to be recognized. When a match between templates occurs, the word can be identified. The various types of speech recognition algorithms include direct comparison of features, Dynamic Time Warping (DTW), and Hidden Markov Modeling (HMM).

Direct comparison   Direct comparison uses speech features directly in the comparison process. For example, in the zero-crossing technique of feature extraction, the counter values obtained for each word are used in the matching process. With LPC filtering, the LPC coefficients are used for comparison.

Dynamic Time Warping (DTW)   Since people do not speak at the same rate, templates of the same word spoken by different people may not match because of the timing difference. DTW is a mathematical process that corrects the timing difference by expanding or compressing the templates during the matching process (Hollingum and Cassford, 1988).

Hidden Markov Modeling (HMM)   HMM is an algorithm that converts the speech input into its own set of codes that number about 200 (Bursky, 1985). These codes can be assembled to create any word in the system's vocabulary. In addition, HMM uses statistics to predict word or sound sequences. The statistical modeling allows the system to differentiate similar sounding words based on context.

Aids for the disabled

With the capability of speech recognition, many applications are possible. Of particular importance are the applications that serve as aids for the disabled. The disabilities that are aided or can be aided by speech recognition include deafness and hearing impairments, motor impairments, and speech impairments.

Deafness and hearing impairments   One application is speech-to-text conversion for television broadcasts and telephone calls. The speech recognizer would convert, in real-time, the messages spoken from the television into readable text displayed on the TV screen. The speech recognition system would need to have continuous-speech/ speaker-independent capabilities and have a large vocabulary. Unfortunately, such a system is not yet available.

A speaker-dependent speech recognition system is available for use with a telephone. The prototype system is called Deafnet (Mills, 1988). Deafnet converts messages typed

by the deaf person into synthesized speech for the hearing
person. If the hearing person has trained Deafnet, then
Deafnet will convert the spoken message into readable text for
the deaf person.

Another application is speech therapy which uses speech
recognition technology. The analysis of speech is displayed
to give visual feedback to a deaf person learning to talk.
This research is being done by George Holland, John Homer, and
Walter Struve at the Ames Laboratory.

Motor impairments   For people with motor impairments,
the obvious application of speech recognition is verbal
control of devices within their environment. Such devices
include lamps, telephones, radios, televisions, thermostats,
page turners, wheel chairs, and personal computers.

Speech impairments   People with serious speech
impairments can benefit from a speech recognition system
trained to understand their utterances. If these utterances
are correlated with the right words, a speech synthesizer can
be used to speak the words they were trying to pronounce
(Mills, 1988). As the degree of impairment increases, the
training time increases and the recognition accuracy
decreases.

Speech Recognition Systems

Many speech recognition systems are commercially
available as listed in Appendix D. Since many of these

systems use specially designed integrated circuits, they are difficult for the experimenter to duplicate. However, simpler speech recognition systems can be built using readily available parts.

Two such speech recognition systems are discussed in the following literature review. These systems provided the inspiration for building a speech recognition system using the computer-controlled digital recorder. Although this system was not completed, the ideas and plans for building the Z8671/DVM-1 system are given in the final section.

Literature review

The first speech recognition system used a Z-80 micro-computer. The system was designed by Mike Rigsby and is explained in his book, Verbal Control with Microcomputers. The second system used a 6502 microcomputer. This system is described in the book, Electrically Hearing: Computer Speech Recognition by John Cater.

The Z-80 system   The speech signal is collected by a microphone and then amplified before being sent to the input port of the processor. The processor converts the speech signal into a squared waveform that is tested by the software to determine the duration of the sound. Based on the variations in the duration of sound, the system can differentiate as many as five different words. Recognition of a word is indicated by five LEDs (light emitting diodes)

connected to the output port of the microcomputer. The shortest duration of sound activates LED #1 and the longest duration turns on LED #5.

The 6502 system   This system uses the 6502 microprocessor developed in 1975 by MOS Technology which is now owned by Commodore Business Machines (Poe, 1983). Although the 6502 is used in this system, the interface circuits can be adapted for use with the Z-80, 8088, or 8080 microprocessors (Cater, 1984).

Hardware   In addition to the 6502, the hardware includes circuits for filtering and amplifying the speech input signal obtained from the microphone. A sample-and-hold circuit samples the processed signal and maintains it until the digitization circuit can make an analog-to-digital conversion. A digital-to-analog converter is also included for reconstructing the original analog speech signal.

Software   The software of the system consists of a machine language subroutine that retrieves the speech data stored in memory, and a speech recognition program written in Microsoft BASIC (Cater, 1984). The BASIC program allows the computer to "learn" as many as 40 words. The program does low-pass, bandpass, and high-pass filtering of the speech data. Zero-crossing detectors and counters for each filter are incorporated in the program. Each vocabulary word is represented by 24 numbers (8 low-pass, 8 bandpass, and 8 high-

pass counter values). The recognition process is accomplished by analyzing an input word and comparing the resulting 24 numbers with all the sets of 24 numbers from words already stored in memory to find the closest match. The recognition process takes about 30 seconds per word making this system slow; however, software and/or hardware changes can decrease the recognition time.

The Z8671/DVM-1 system

Since the DVM-1 is able to digitize the speech signal and store it in memory, most of the hardware for a speech recognition system is already available. To complete the system, the speech data must be accessed from the D-RAM, and speech recognition software must be written.

Speech data acquisition   Under control of the T6668 speech microprocessor, speech data travels to and from the D-RAM in serial fashion. One possible method of data acquisition is to tap into the D-RAM serial output line and connect it to the port 3 input line of the Z8671 which is part of an asynchronous receiver system. The receiver also consists of a receiver buffer and a serial-in, parallel-out shift register (Zilog, Inc., 1984). The received data must have a start bit, eight data bits, and at least one stop bit. The rate at which the data is read is controlled by the counter/timer register and the prescaler register of the

Z8671. This rate must be coordinated with the bit rate set by the T6668.

If the bit rate of the transmitter and receiver vary by more than +/- 5%, the D-RAM may have to be controlled by a D-RAM controller instead of the T6668 during speech data acquisition (Metzger, 1989). The easiest way to access the D-RAM pins for connection to the controller is to remove one of the D-RAMs from its socket mounted to the DVM-1 board.

Another approach to reading the serial data is to use the Z-80 serial input/output controller (Z-80 SIO). The Z-80 SIO connects directly to the Z8671 and provides two independent channels for asynchronous or synchronous serial communication (Uffenbeck, 1985).

Software   The software is needed to control the movement of speech data and also to perform the speech recognition. The vocabulary of the system is stored in the D-RAM. Once the word to be recognized is digitized, its speech data are retrieved and stored in the RAM of the Z8671. The set of speech data representing a vocabulary word is then retrieved and stored in RAM. A speech recognition algorithm then compares the two sets of data. This process is repeated for each vocabulary word. The algorithm recognizes the word based on the best match between sets of speech data. Undoubtedly, better hardware/software techniques are possible for extracting features and performing speech recognition.

RESULTS AND RECOMMENDATIONS

The DVM-1 recorder was successfully interfaced to the MC-1Z computer. In addition to its function in a speech recognition system, the computer-controlled recorder offers many more possible applications. One major application involves interfacing the system to monitoring equipment so that the measured parameter can be announced in addition to being visually displayed. For example, the MC-1Z/DVM-1 system could be interfaced to a heart rate monitor and the appropriate software written to allow the system to announce the heart rate.

Although the software written for the computer-controlled recorder is functional, much more sophisticated software can be developed. One improvement for recording in the Direct mode is a visual display indicating which sections of memory are available for recording. The sections of memory could be listed by start and stop address and by how much time (seconds) is available for each recording bit rate. The key to improved software is to make the recording process as easy and convenient as possible for the user, but still provide plenty of flexibility in using the recorder.

One final important consideration is maintaining voice data in the D-RAM. Some applications may depend on the system working even when power is interrupted. A battery-backup system becomes necessary in this situation. Another reason

for having battery-backup is to save power. When the sytem is not recording or reproducing, the batteries power the T6668 and D-RAM and prevent loss of voice data until normal power is restored.

The MC-1Z/DVM-1 speech recognition system was not implemented. The main obstacle was accessing the voice data from the D-RAM. Without access to the data, the speech recognition algorithm has nothing to analyze. Once the voice data are accessed, a speech recognition system is possible using the computer-controlled recorder.

While researching speech recognition, it was learned that one of the most important aspects of a successful speech recognition system is the speech recognition algorithm. Although delta modulation (method of speech encoding used in the DVM-1) is rarely used in speech recognition systems (Cater, 1984), a good project would be to develop speech recognition software that can accept delta encoded speech input. No matter what speech recognition approach is used, the software should be programmed in machine language to make the recognition process as fast as possible.

With limited memory, the MC-1Z/DVM-1 system probably cannot support speaker-independent speech recognition; however, many applications exist for a speaker-dependent system. These applications include verbal control of the MC-1Z and the many other previously mentioned applications

that can help people with hearing, speech, and motor
impairments.

BIBLIOGRAPHY

Basicon, Inc. 1984. MC-1Z Microcontroller. Basicon, Inc., Portland, Oregon.

Bristow, Geoff. 1984. Electronic speech synthesis: Techniques and applications. McGraw-Hill Book Company, New York. 346 pages.

Bursky, Dave. 1985. Speech recognition builds its vocabulary to handle more tasks. Electronic Design 33(9):113-124.

Cater, John P. 1984. Electronically hearing: Computer speech recognition. Howard W. Sams and Co., Inc., Indianapolis, Indiana. 263 pages.

Feher, K. 1987. Telecommunications measurements, analysis, and instrumentation. Prentice-Hall, Inc., Englewood Cliffs, New Jersey. 412 pages.

Holland, George E. Undated. Voice systems: A visual display for interpreting and teaching speech. Ames Laboratory, Iowa State University.

Hollingum, Jack and Graham Cassford. 1988. Speech technology at work. IFS Ltd., Kempston, Bedford, UK. 158 pages.

Lesea, Austin and Rodnay Zaks. 1979. Microprocessor interfacing techniques. Sybex, Berkeley, California. 456 pages.

Metzger, Daniel L. 1989. A practical approach to hardware, software, troubleshooting, and interfacing. Prentice Hall, Inc., Englewood Cliffs, New Jersey. 679 pages.

Mills, Robert. 1988. A survey of speech technology used by people with disabilities. Speech Technology 4(3):56-60.

The 1989 speech technology buyer's guide. 1989. Speech Technology 4(4):85-114.

Poe, Elmer. 1983. The microprocessor handbook. Howard W. Sams and Co., Inc., Indianapolis, Indiana. 236 pages.

Poulton, A. S. 1983. Microcomputer speech synthesis and recognition. Sigma Technical Press, Wilmslow, Cheshire, U.K. 193 pages.

Rigsby, Mike. 1982. Verbal control with microcomputers. Tab Books, Inc., Blue Ridge Summit, Pennsylvania. 304 pages.

Texas Instruments, Inc. 1984. MOS memory databook. Texas Instruments, Inc., Dallas, Texas.

Toshiba America, Inc. 1988. Speech devices databook. Toshiba America, Inc., Irvine, California.

Uffenbeck, John. 1985. Microcomputers and microprocessors: The 8080, 8085, and Z-80 programming, interfacing, and troubleshooting. Prentice-Hall, Inc., Englewood Cliffs, New Jersey. 670 pages.

Wallich, Paul. 1987. Putting speech recognizers to work. IEEE Spectrum (USA) 24(4):55-57.

Zilog, Inc. 1981. Z8671 single-chip BASIC interpreter: BASIC/DEBUG software reference manual. Zilog publication ref. no.: 03-3149-02. Zilog, Inc., 1315 Dell Ave., Campbell, California.

Zilog, Inc. 1984. Z8 microcomputer: Technical manual. Zilog publication ref. no.: 03-3047-02. Zilog, Inc., 1315 Dell Ave., Campbell, California.

ACKNOWLEDGEMENTS

I would like to thank my major professor, Dr. Curran S. Swift, for his guidance, technical expertise, and suggestions. On a personal basis, I want to express my appreciation for his consideration, pleasant disposition, and constant willingness to help at any time.

I would also like to thank Drs. Dave Carlson, Chester Comstock, and Mary Helen Greer for serving on my graduate committee. I am particulary indebted to Dr. Greer for her thoughtful and ambitious help as an instructor and head of the department.

In addition, I would like to thank the faculty, staff, and students of the Biomedical Engineering Department who made my experience at Iowa State University a memorable and pleasant one.

Finally, and most importantly, I want to express my deepest gratitude to my parents, Raymond and Nancy Fagan, for their guidance and support. Without their hard work, dedication, and sacrifice, I would not have the opportunities that I have today.

APPENDIX A:

T6668 PIN CONNECTIONS
(Toshiba America, Inc., 1988)

60 PIN FLAT PACKAGE

Top view pin connections:

Top pins (45–31): 256K, TEST, X_OUT, X_IN, EOS, RD, WR, N.C, CE, NC, D7, D6, D5, D4, D3

Left side pins (46–60):
- 46 — WE
- 47 — RAS
- 48 — D_IN
- 49 — D_OUT
- 50 — A0
- 51 — A1
- 52 — N.C
- 53 — V_DD
- 54 — A2
- 55 — A3
- 56 — A4
- 57 — A5
- 58 — A6
- 59 — A7
- 60 — A8

Right side pins (30–16):
- 30 — N.C
- 29 — D2
- 28 — D1
- 27 — D0
- 26 — CPUM
- 25 — ACL
- 24 — N.C
- 23 — V_DD
- 22 — ADI
- 21 — DAO
- 20 — FILIN
- 19 — FILOUT
- 18 — TS2
- 17 — TS1
- 16 — MICIN

Bottom pins (1–15): CAS1, CAS2, N.C, CAS3, CAS4, EXT, M1, M2, V_SS1, V_SS2, C_Bref, V_ref, MICOUT, C2, C1

* NC --- No connection

APPENDIX B:

A SAMPLE PROGRAM FOR RECORDING IN THE LABEL/INDEX MODE

```
10 REM Recording in the Label/Index mode

20 REM Assigning variables to port locations

21 A=%B800:   B=A+1

30 REM Setting ports A and B for output

31 @%B803=%80

40 REM Resetting the address counter

41 @B=7:   @B=15

50 REM Writing the REC command

51 @A=128:   @B=10:   @B=15

60 REM Writing the CNDT command

61 @A=83:   @B=10:   @B=15

70 REM Writing the LABEL command

71 PRINT "ENTER PHRASE NUMBER(0, 1,..., OR 15) ";: INPUT P

72 IF P>15 THEN GOTO 71

73 IF P<0 THEN GOTO 71

74 @A=96+P:   @B=10:   @B=15

80 REM Writing the STOP command

81 PRINT "PRESS S TO STOP RECORDING ";: W=USR(84)

82 IF W<>83 THEN GOTO 81

83 @A=32:   @B=10:   @B=15
```

APPENDIX C:

A SAMPLE PROGRAM FOR RECORDING IN THE DIRECT MODE

```
10 REM Recording in the Direct mode

20 REM Assigning variables to port locations

21 A=%B800:   B=A+1

30 REM Setting ports A and B for output

31 @%B803=%80

40 REM Resetting the address counter

41 @B=7:   @B=15

50 REM Writing the REC command

51 @A=128:   @B=10:   @B=15

60 REM Writing the CNDT command

61 @A=83:   @B=10:   @B=15

70 REM Writing the ADLD1 command

71 PRINT "START ADDRESS"

72 PRINT "ENTER BYTE #1, #2, AND #3 IN %FF FORM";: INPUT N,O,P

73 @A=N:   @B=10:   @B=15
74 @A=O:   @B=10:   @B=15
75 @A=P:   @B=10:   @B=15

80 REM Writing the ADLD2 command

81 PRINT "STOP ADDRESS"

82 PRINT "ENTER BYTE #1, #2, AND #3 IN %FF FORM";: INPUT Q,R,S

83 @A=Q:   @B=10:   @B=15
84 @A=R:   @B=10:   @B=15
85 @A=S:   @B=10:   @B=15

90 REM Writing the START command

91 @A=16:   @B=10:   @B=15
```

APPENDIX D:

MANUFACTURERS OF SPEECH RECOGNITION SYSTEMS
(The 1989 speech, 1989, Poulton, 1983, Wallich, 1987)

AT&T Conversant Systems
6200 East Broad Street
Columbus, OH 43213
(614) 860-2000

    speaker-independent,
    isolated word

    vocabulary: ?[1]

Animated Voice Corporation

    speaker-independent,
    continuous-speech

    vocabulary: ?

Articulate Systems, Inc.

    speaker-independent,
    isolated-word

    vocabulary: 1000 words

Audec Corporation
299 Market Street
Saddle Brook, NJ 07662
(201) 368-3848

    speaker-independent,
    isolated-word

    vocabulary: 144 words

Automated Call Processing

    speaker-independent,
    isolated-word

    vocabulary: 16 words

Automation Electronics Corp.

    speaker-dependent,
    ?

    vocabulary: ?

Axxon Voice Products

    speaker-dependent,
    isolated-word

    vocabulary: 200 words

Berkeley Speech Technologies
2409 Telegraph Avenue
Berkeley, CA 94704
(415) 841-5083

    speaker-dependent,
    isolated word

    vocabulary: ?

---

[1]Information not available.

California Medical Software   speaker-dependent,
                                              isolated-word

                                              vocabulary: 500+ words

Computer Consoles, Inc.   speaker-independent,
97 Humboldt Street   isolated-word
Rochester, NY 14609
(716) 482-5000   vocabulary: 50 words

Covox, Inc.   speaker-dependent,
                                              isolated-word

                                              vocabulary: 64 words

Dictaphone Corporation   speaker-dependent,
120 Old Post Road   isolated-word
Rye, NY 10580
(914) 967-7300   vocabulary: ?

Digital Equipment Corporation   speaker-dependent,
146 Main Street   isolated-word
Maynard, MA 01754
                                              vocabulary: ?

Digitron Telecommunications   speaker-independent,
                                              continuous-speech

                                              vocabulary: ?

Dragon Systems, Inc.   speaker-dependent,
173 Highland Street   isolated-word
West Newton, MA 02165
(617) 527-0372   vocabulary: 5K-60K words

Fujitsu America, Inc.   speaker-dependent,
                                              isolated-word

                                              vocabulary: 4000 words

Gralin Associates, Inc.   speaker-independent,
                                              isolated-word

                                              vocabulary: 20 words

Hear Say Inc.   speaker-independent,
1845 74th Street   isolated-word
Brooklyn, NY 11204
                                              vocabulary: 64 words

IBM
590 Madison Avenue
New York, NY 10022
(212) 735-7000

speaker-dependent,
isolated-word

vocabulary: 64 words

Intel Corporation
3065 Bowers Avenue
Santa Clara, CA 95051

speaker-dependent,
isolated-word

vocabulary: 200 words

Intellisystems, Inc.

speaker-independent,
isolated-word

vocabulary: 14 words

International Voice Products, Inc.

speaker-dependent,
isolated-word

vocabulary: 400 words

Interstate Voice Products
1849 West Sequoia Avenue
Orange, CA 92668
(714) 937-9010

speaker-dependent,
continuous-speech

vocabulary: 100 words

Intervoice, Inc.
1850 N. Greenville Ave.
Richardson, TX 75081
(214) 669-3988

speaker-independent,
isolated-word

vocabulary: 16 words

IOCS, Inc.
400 Totten Pond Road
Waltham, MA 02254
(202) 879-7000

speaker-independent,
isolated-word

vocabulary: 5400 words

Kurzweil Applied Intelligence
411 Waverley Oaks Road
Waltham, MA 02154
(617) 893-5151

speaker-dependent,
isolated-word

vocabulary: 10K words

Mimic, Inc.
P.O. Box 921
Acton, MA 01720

speaker-dependent,
isolated word

vocabulary: 79 words

NEC America, Inc.
532 Broad Hollow Road
Melville, NY 11747
(516) 752-9700

speaker-dependent,
isolated word

vocabulary: 250 words

Scott Instruments Corporation 1111 Willow Springs Drive Denton, Texas 76201 (817) 387-9514	speaker-independent, isolated-word vocabulary: 150 words
Speech Systems, Inc. 18356 Oxnard Street Tarzana, CA 91356 (818) 881-0885	speaker-independent, continuous-speech vocabulary: 36K words
Tandy/Intertan Canada	speaker-dependent, isolated-word vocabulary: 128K words
Technologia Systems, Ltd.	speaker-dependent, isolated-word and continuous-speech vocabulary: 20K+ words
Telecorp Systems, Inc.	speaker-independent, continuous-speech vocabulary: ?
Telsis, Ltd.	speaker-dependent, isolated-word vocabulary: 16 words
Texas Instruments, Inc. P.O. Box 401560 Dallas, TX 75240 (214) 680-5096	speaker-dependent, isolated-word vocabulary: 1000 words
Toshiba America, Inc. 375 Park Avenue New York, NY 10152 (212) 308-2040	speaker-dependent, isolated-word vocabulary: 40 words
Transyn, Inc.	speaker-dependent, isolated-word vocabulary: 600 words
Verbex Voice Systems Two Oak Park Bedford, MA 01730 (617) 275-5160	speaker-dependent, continuous-speech vocabulary: 800 words

Vocollect, Inc.

speaker-dependent,
isolated-word

vocabulary: 1000 words

Voice Connection

Irvine, CA

speaker-dependent,
isolated-word

vocabulary: 500 words

Voice Control Systems
14140 Midway Road
Dallas, TX 75244
(214) 386-0300

speaker-independent,
isolated-word

vocabulary: 40 words

Voicetek Corporation

speaker-independent,
continous-speech

vocabulary: 12 words

Votan
4487 Technology Drive
Fremont, CA 94538
(415) 490-7600

speaker-independent,
isolated-word and
continuous-speech
vocabulary: 13 words

Westinghouse Voice Systems
701 Rodi Road
Pittsburgh, PA 15235
(412) 825-3500

speaker-independent,
continuous-speech

vocabulary: ?