MAY 21 1993

silicon systems®
A TDK Group Company

SSI 78A263A
Phoneme Speech Synthesizer

February 1993

DESCRIPTION

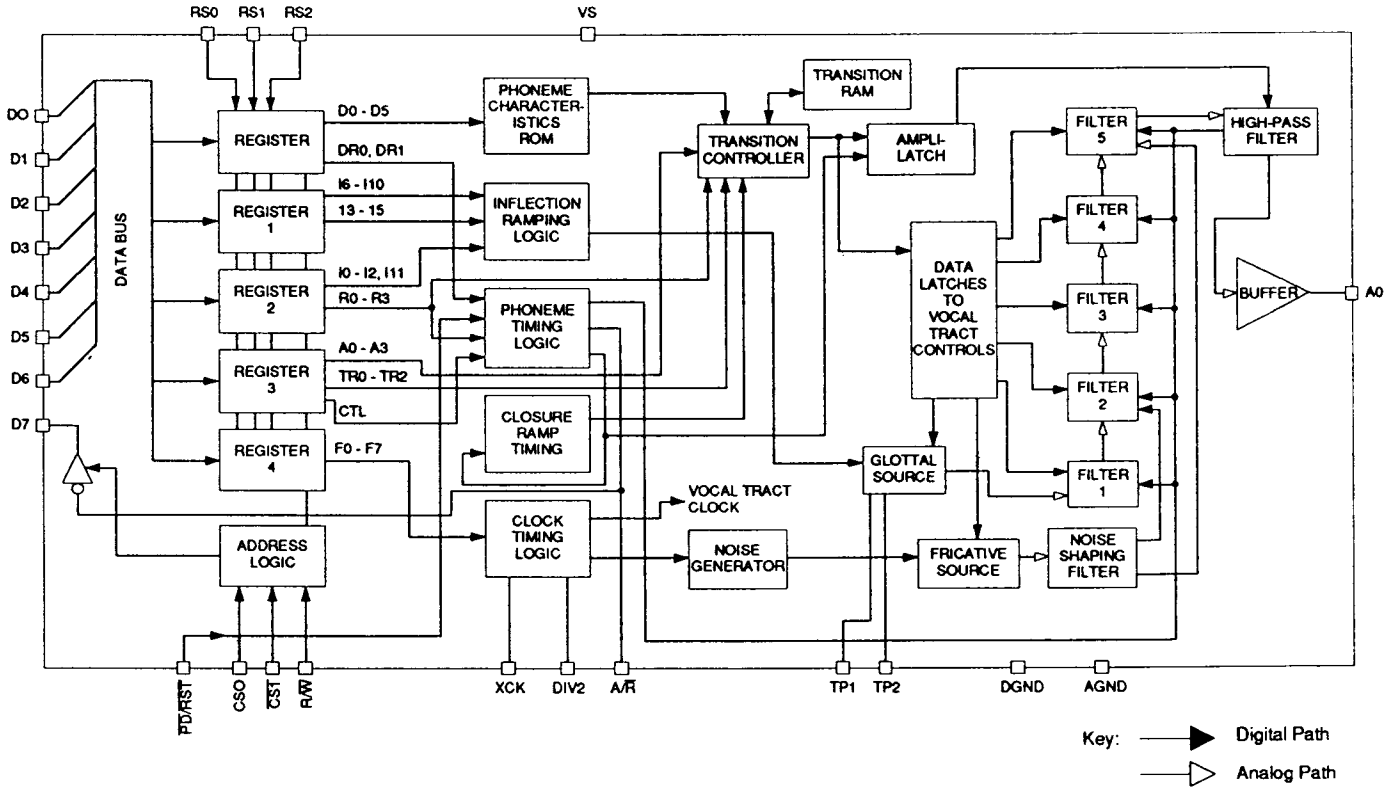
The SSI 78A263A is a versatile, high-quality, phoneme-based speech synthesizer circuit contained in a single, monolithic CMOS integrated circuit. It is designed to produce an audio output of unlimited vocabulary, music and sound effects at an extremely low data input rate.

Speech is synthesized by combining phonemes, the building blocks of speech, in an appropriate sequence. The SSI 78A263A contains five, eight-bit registers that allow software control of speech rate, pitch, pitch movement rate, amplitude, articulation rate, vocal tract filter response, and phoneme selection and duration.

FEATURES

- Single, low-power CMOS integrated circuit
- 5 Volt supply
- Extremely low data rate
- 8-bit bus compatible with selectable handshaking modes
- Non-dedicated speech, ideal for text-to-speech programming
- Programmable and hard Power-down/Reset mode
- Switched-capacitor-filter technology

BLOCK DIAGRAM



SSI 78A263A
Phoneme
Speech Synthesizer

FUNCTIONAL DESCRIPTION

This short description is intended to provide SSI 78A263A feature and capability information only.

PRODUCTION OF SPEECH

To produce different speech phonemes (sounds) the SSI 78A263A uses a model of the human vocal tract. Within the device this analog tract is modeled with five cascaded programmable low pass filter sections. The filter sections are programmed internally by a digital controller. Either a glottal (pitch) or a pseudo-random noise source is used to excite the vocal tract, depending on whether a voiced or non-voiced phoneme is selected. During speech production the phonemes will typically last between 25 and 100 ms.

SPEECH ATTRIBUTE REGISTERS

Speech is produced by programming speech attribute (characteristic) data into five, eight-bit registers. These internal registers allow selection of phonemes and speech characteristics. Refer to the Register Input Formats for the functional allocations.

DEVICE RESPONSE TO ATTRIBUTE REGISTER DATA

The SSI 78A263A has two general classes of attribute data: "control" data (speech rate, filter frequency, phoneme articulation rate, phoneme duration, immediate inflection setting, and inflection movement rate) and "target" data (phoneme selection, audio amplitude, and transitioned inflection). The SSI 78A263A responds immediately upon loading "control" data; upon loading "target" data the device will begin to move towards that target at the prescribed transition rates. This fully internal linear transitioning between target values, done in a manner as is found in normal speech, is a key factor in reducing control data rate without sacrificing speech quality.

ATTRIBUTE REGISTER WRITING

The eight-bit data bus D7-D0 loads the particular attribute register selected by the three bit address bus RS2-RS0. To write the data, R/$\overline{W}$ (Read/Write), CS0 (Chip Select 0), and $\overline{CS1}$ pins must first be in the 0, 1, 0 state, respectively. The data is then written when at least one of these pins changes state. Refer to the Write Timing Diagram. Writing is accomplished by changing preferably CS0 or $\overline{CS1}$. Following device power up, nominal values should be loaded into the attribute registers as described below.

APPROXIMATE DATA TRANSFER RATE

For speech production using the SSI 78A263A, the actual data rate depends on the amount of speech attribute manipulation. For example, the production of monotonic speech, where phoneme and duration are the only attribute manipulations, requires a data rate of less than 100 bit/s. A higher data rate of about 500 bit/s is required for high quality speech due to the associated full attribute manipulation.

SELECTABLE OPERATION MODES

The state of the Duration/Phoneme Register bits DR1 and DR0 determine the operating mode of the device when the control bit (CTL) is changed from a logic one to a logic zero. The four modes of operation include choice of timing response between "frame" or "phoneme" timing (as explained below), transitioned or immediate inflection response, and setting the A/$\overline{R}$ (Acknowledge/Request Not) pin active or disabled. Refer to the Mode Selection Chart.

PHONEME SELECTION

The SSI 78A263A can produce the 64 phonemes listed on the Phoneme Chart. Bits P5-P0 are used for phoneme selection. The relative phoneme duration is set by bits DR1 and DR0.

PHONEME ARTICULATION ADJUSTMENT

A particular phoneme is produced by the combination of vocal-tract low-pass filter settings, excitation source type, and source amplitude. When a new phoneme is selected, the device performs a linear transition to the new set of characteristics. The rate of this transition is controlled by the articulation setting, bits TR2-TR0. This rate is relative in that articulation is not affected by speech rate bits R3-R0. A typical articulation register setting is "5."

PROGRAMMING INFLECTION (PITCH)

When the SSI 78A263A is in the mode of immediate inflection, bits I11-I0 provide immediate adjustment with seven octaves of pitch on a even tempered scale. With the device in the transitioned inflection mode, bits I10-I6 select the target pitch and bits I5-I3 determine the inflection rate of change. Bits I11, I2, I1, and I0 always provide immediate adjustment. A typical value used for speech production is 90 Hz where:

$$\text{InflectionFrequency} = \frac{\text{XCK frequency}}{8 \times (4096 - I)}$$

I = decimal equivalent of Inflection Register setting

Filter Frequency Setting

Data bits FF7-FF0 set the clock frequency for the switched-capacitor vocal tract filters. This determines overall filter frequency response. Inflection pitch is not affected by these bits. Typically this is set to give a clock frequency of about 20 kHz (see formula below), but may be manipulated to fine-tune speech quality or to change "voice type"; bass, baritone, etc.

$$\text{Filter Frequency} = \frac{\text{XCK frequency}}{2\,(256-\text{FF})}$$

FF = decimal equivalent to the Filter Frequency Register setting.

Speech Rate

Rate of speech is controlled by bits R3-R0, the Speech Rate Register. In Frame Timing mode new attribute data is requested at the end of a "frame" where:

$$\text{Frame Duration} = \frac{4096 \times (16-\text{R})}{\text{XCK frequency}}$$

R = decimal equivalent of Rate Register setting. In the Phoneme Timing Mode the frame duration is modified by the phoneme duration bits DR1 and DR0 where:

Phoneme Duration = (Frame Duration) X (4-D)

D = decimal equivalent of Duration Register setting.

All internal attribute transitioning is performed relative to the Speech Rate Register setting. Speech rate does not affect inflection or filter frequency. A typical rate setting is hexadecimal "A."

Amplitude Adjustment

The overall audio output level is set with register bits A3-A0. Since each phoneme has a preset amplitude relative to other phonemes, it is not necessary to program the amplitude of each phoneme; however, amplitude changes may be used to enhance the speech quality and add emphasis. Amplitude is transitioned linearly at a rate dependent on the phoneme duration setting. A typical amplitude setting is hexadecimal "C."

Control Bit and Power Down Mode

Setting the control bit (CTL) to a logic one puts the device into Power Down mode, a sort of "standby." This bit is also set high when the $\overline{PD}/\overline{RST}$ pin is brought low and also upon power up. The Power Down mode turns off the excitation sources and analog circuits to reduce power consumption, but maintains the present register settings. Upon a control bit logic one-to-zero transition, the present settings of DR1 and DR0 determine the operation mode as described above.

Register Reading

Device pin D7 becomes an output, as the inverted state of A/$\overline{R}$, when the device is put into Read (R/$\overline{W}$ is a logic 1 and the chip is selected, $\overline{CS1}$ = 0, CS0 = 1). Refer to the Read Timing Diagram. The register address bits are ignored.

Time Base

Many different time bases may be utilized (see external clock input specification). It is desirable to establish a stable crystal controlled time base from 800 to 1000 kHz when DIV2 is set low, or twice the frequency when DIV2 is set high. A good time base can be easily accomplished with an inexpensive colorburst 3.5795 MHz crystal in conjunction with a divide-by-two circuit. The actual device timing and output frequencies are directly related to the time base frequency used.

Microprocessor Interfacing

Either the A/R line, or D7 as an output, are used as an interrupt to indicate when the duration of a frame or phoneme has been exceeded. No detectable degradation to speech quality results when several milliseconds occur between data request and load.

PIN DESCRIPTION

NAME	TYPE	DESCRIPTION
AO	O	Analog Audio Output-biased @ VDD/2 requires an external audio amp for speaker drive.
AGND	-	Analog Ground.
TP1	-	Do not use.
A/$\overline{R}$	O	Acknowledge/Request Not – open collector output changes from high to low level after phoneme is generated. May be used as an interrupt request for new phoneme data (see D7 also).
TP2	-	Do not use.
RS2	I	Register Select Input – used to select one of five internal registers in conjunction with RS1 and RS0.
RS1	I	Register Select (see RS2 pin).
RS0	I	Register Select (See RS2 pin).
D0	I	LSB of 8-bit data bus – input only.
D1	I	Data Input (only).
D2	I	Data Input (only).
DGND	-	Digital Ground.
D3	I	Data Input (only).
D4	I	Data Input (only).
D5	I	Data Input (only).
D6	I	Data Input (only).
D7	I/O	MSB of 8-bit data bus. Bi-directional, inverse of A/$\overline{R}$ when read is high.
$\overline{PD/RST}$	I	Power Down Control Input – Silences audio output and retains DC bias without disturbing register contents. Disables A/$\overline{R}$ output.
CS0	I	Chip Select Input.
$\overline{CS1}$	I	Chip Select Input.
R/$\overline{W}$	I	Read/Write Control Input – Write is active low for loading internal registers. Read is active high but enables D7 only.
XCK	I	Clock Input ($\simeq$1 or 2 MHz)
DIV2	O	Clock Divide by Two – used when external clock is $\simeq$ 2 MHz.
VDD	-	Positive Voltage Supply.

ELECTRICAL SPECIFICATIONS

Unless otherwise specified, 4.5≤ VDD ≤ 5.5; 0°C ≤ Ta ≤ 70°C
1.50 MHz ≤ XCK frequency ≤ 2.0 MHz, when XCK/2 = logic 1 or 0.75 MHz ≤ XCK frequency ≤ 1.0 MHz, when XCK/2 = logic 0.

ABSOLUTE MAXIMUM RATINGS (Operation above maximum ratings may damage the device.)

PARAMETER		RATING
Supply Temperature	VDD - VSS	7.0V
Input Voltage	VIN	-0.5 to VDD + 0.5V
D.C. Current at Inputs	IINM	±1.0 mA
Storage Temperature	TS	-55 to +125°C
Operating Temperature	TA	0 to 70°C
Power Dissipation	Pd	500 mW

POWER SUPPLY

PARAMETER	CONDITIONS	MIN	NOM	MAX	UNIT
Supply Current	$\overline{PD}/\overline{RST}$ = 1, CTL = 0		8	20	mA
Supply Current	$\overline{PD}/\overline{RST}$ = 0, CTL = 1		7	18	mA

AUDIO OUTPUT

Output Level	AW phoneme; RL = 50 kΩ to GND through 1 µF cap.	0.28 VDD	0.37 VDD	0.50 VDD	Vpp
DC Output Offset		0.5 VDD	0.6 VDD	0.7 VDD	V
Resistive Loading	AC coupled to AO to GND	10			kΩ
Capacitive Loading	To GND to ensure Stable A			100	pF

BUS CONROL INPUTS, DATA INPUTS (RS0, RS1, RS2, CS0, $\overline{CS1}$, D0-D7 PD/$\overline{RST}$)

Input High Voltage	VIH		VSS +2.4		VDD +0.3	VDC
Input Low Voltage	VIL		–0.3		+0.8	VDC
Input Leakage Current	IIN	VIN = 0 to VDD			5	µA
Input Capacitance	CIN	VIN = 0 Ta = 25°C measured at f = 1.0 MHz			10	pF
Input Capacitance, D7 Input	CIN (D7)				20	pF
Input Current, D7 in TRI-State "OFF" State	IN(TS)	VIN = 0.4 to 2.4 V		2.0	5.0	µA

D7 OUTPUT

D7 Output Low Voltage	VOL(D7)	ILoad = 0.4 mA into D7			0.4	VDC
D7 Output High Voltage	VOH (D7)	ILoad = 205 µA out of D7			VDD -2.0	VDC

ELECTRICAL SPECIFICATIONS (continued)

A/$\overline{R}$ OUTPUT

PARAMETER		CONDITIONS	MIN	NOM	MAX	UNIT
Output Low Voltage	IOL (A/$\overline{R}$)	IL = 3.2 mA into A/$\overline{R}$			0.4	VDC
Output High Leakage Current	IL (A/$\overline{R}$)	VOut = 0 to VDD			10	µA
Output Capacitance	COut (A/$\overline{R}$)	VOut = 0 VDC Ta = 25°C f = 1.0 MHz			15	pF

DIV2 INPUT

Input Low Voltage	VIL (DIV2)		-0.3		0.2VDD	V
Input High Voltage	VIH (DIV2)		0.8VDD		VDD+0.3	V
Input Leakage		VIN = 0 to VDD			5	µA

XCLK

Input Low Voltage	VIH (IC)		–0.3		+0.8	V
Input High Voltage	VIH(IC)		2.4		VDD+0.3	V
Input Current	IIN(C)	VIN = 0.0 to VDD			5	µA
Input Capacitance	CIN(C)				10	pF
Duty Cycle	D(XCLK)		0.4		0.6	–

TABLE 1: Phoneme Chart

Hex Code*	Phoneme Symbol	Example Word (or Usage)
00	PA	(pause)
01	E	MEET
02	E1	BENT
03	Y	BEFORE
04	YI	YEAR
05	AY	PLEASE
06	IE	ANY
07	I	SIX
08	A	MADE
09	AI	CARE
0A	EH	NEST
0B	EH1	BELT
0C	AE	DAD
0D	AE1	AFTER
0E	AH	GOT
0F	AH1	FATHER
10	AW	OFFICE
11	O	STORE
12	OU	BOAT
13	OO	LOOK
14	IU	YOU
15	IU1	COULD
16	U	TUNE
17	U1	CARTOON
18	UH	WONDER
19	UH1	LOVE
1A	UH2	WHAT
1B	UH3	NUT
1C	ER	BIRD
1D	R	ROOF
1E	R1	RUG
1F	R2	MUTTER (German)
20	L	LIFT
21	L1	PLAY
22	LF	FALL (final)
23	W	WATER
24	B	BAG
25	D	PAID
26	KV	TAG (glottal stop)
27	P	PEN
28	T	TART
29	K	KIT
2A	HV	(hold vocal)
2B	HVC	(hold vocal closure)
2C	HF	HEART
2D	HFC	(hold fricative closure)
2E	HN	(hold nasal)
2F	Z	ZERO
30	S	SAME
31	J	MEASURE
32	SCH	SHIP
33	V	VERY
34	F	FOUR
35	THV	THERE
36	TH	WITH
37	M	MORE
38	N	NINE
39	NG	RANG
3A	:A	MARCHEN (German)
3B	:OH	LOWE (French)
3C	:U	FUNF (German)
3D	:UH	MENU (French)
3E	E2	BITTE (German)
3F	LB	LUBE

*Hex codes shown with DR0, DR1 = 0 (longest duration)

TABLE 2: Register Input Formats

Register Address			Register Name	Bus Input Bit Position							
RS2	RS1	RS0		D7	D6	D5	D4	D3	D2	D1	D0
LO	LO	LO	Duration/Phoneme(DR/P)	DR1	DR0	P5	P4	P3	P2	P1	P0
LO	LO	HI	Inflection (I)	I10	I9	I8	I7	I6	I5	I4	I3
LO	HI	LO	Rate/Inflection (R/I)	R3	R2	R1	R0	I11	I2	I1	I0
LO	HI	HI	Control/Articulation/ Amplitude (C/A/A)	CTL	T2	T1	T0	A3	A2	A1	A0
HI	X	X	Filter Frequency (F)	F7	F6	F5	F4	F3	F2	F1	F0

DR1, DR0 ... Define the phoneme duration.
P5 - P0 ..... Address the phoneme required.
I11 - I0 ..... Define inflection target frequencies and rate of change.
R3 - R0 ..... Define the rate or speed of speech.
CTL ..... Define the mode of A/$\overline{R}$ response in conjunction with DR1 and DR0.
..... Also directly set by PD/$\overline{RST}$.
T2 - T0 ..... Define the rate of movement of the format position for articulation purposes.
A3 - A0 ..... Define the amplitude of the output audio.
F7 - F0 ..... Define the frequency of all vocal tract filters.

TABLE 3: Mode Selection Chart

DR1	DR0	'CTL' BIT	Function
HI	HI	HI - LO	A/$\overline{R}$ active; phoneme timing response; transitioned inflection (most commonly used mode)
HI	LO	HI - LO	A/$\overline{R}$ active; phoneme timing response; immediate inflection
LO	HI	HI - LO	A/$\overline{R}$ active; frame timing response; immediate inflection
LO	LO	HI - LO	Disables A/$\overline{R}$ output only; does not change previous A/$\overline{R}$ response

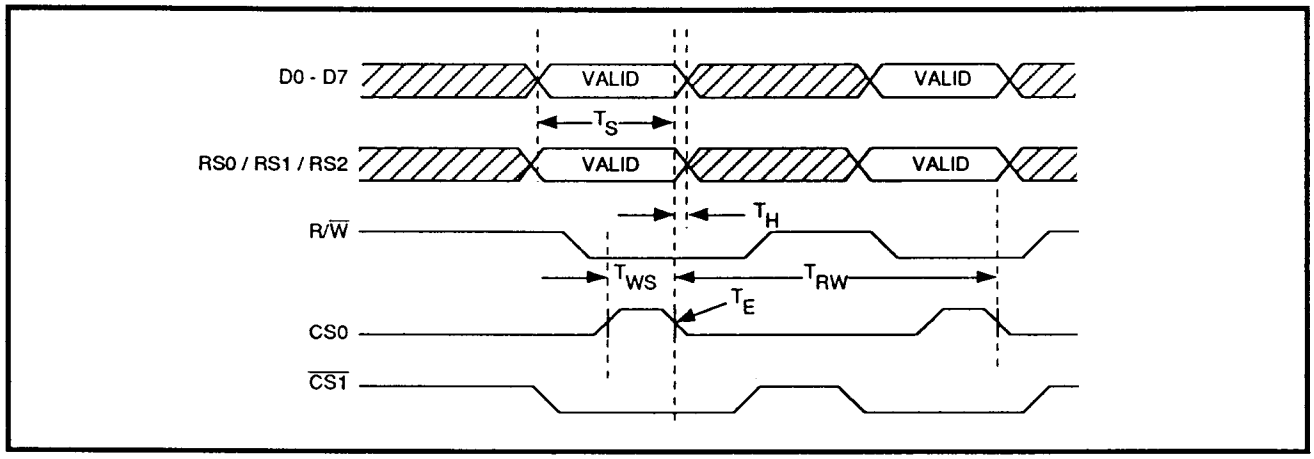


FIGURE 1: Write Timing Diagram

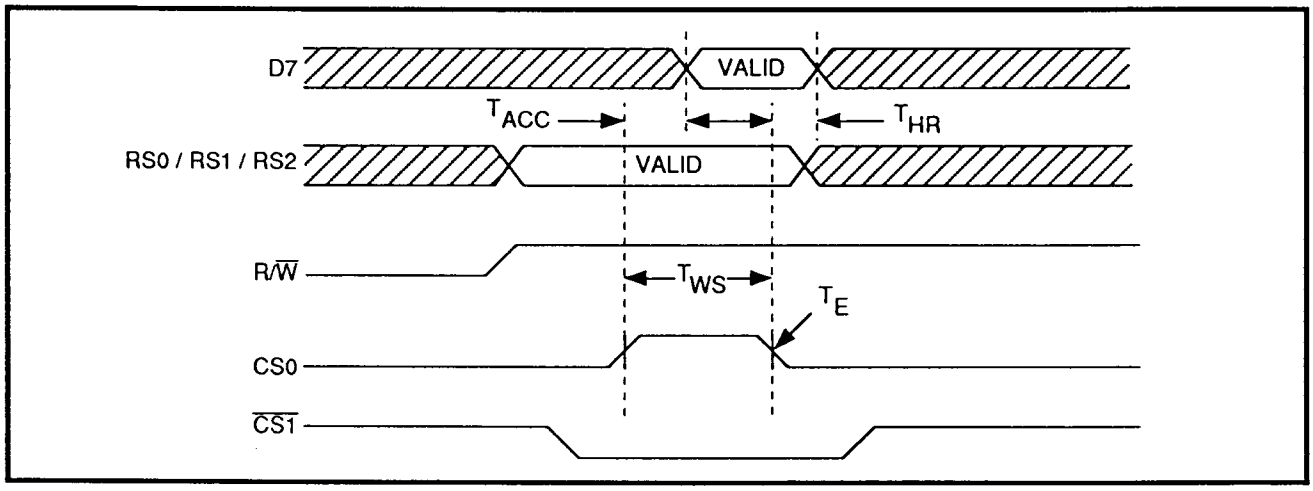


FIGURE 2: Read Timing Diagram

*Valid data latched on first rise or fall of R/$\overline{W}$, CS0, or $\overline{CS1}$ into inactive

TIMING CHARACTERISTICS (VDD = 4.5 to 5.5V, Ta = -40 to + 85°C)

PARAMETER		CONDITIONS	MIN	NOM	MAX	UNIT
Data Setup Time	TS	see note 2	120			ns
Data Hold Time	TH	see note 2	10			ns
Strobe Width	TWS		200			ns
Read/Write Cycle Time	TRW	see note 1	2.25			µs
Rise/Fall Time	TE				100	ns
D7 Output Access Time	TACC				180	ns
D7 Output Hold Time	THR				180	ns

Notes: 1. Based on color burst frequency.

2. Timing relative to deselect by either CS0, CS1, or R/W changing.

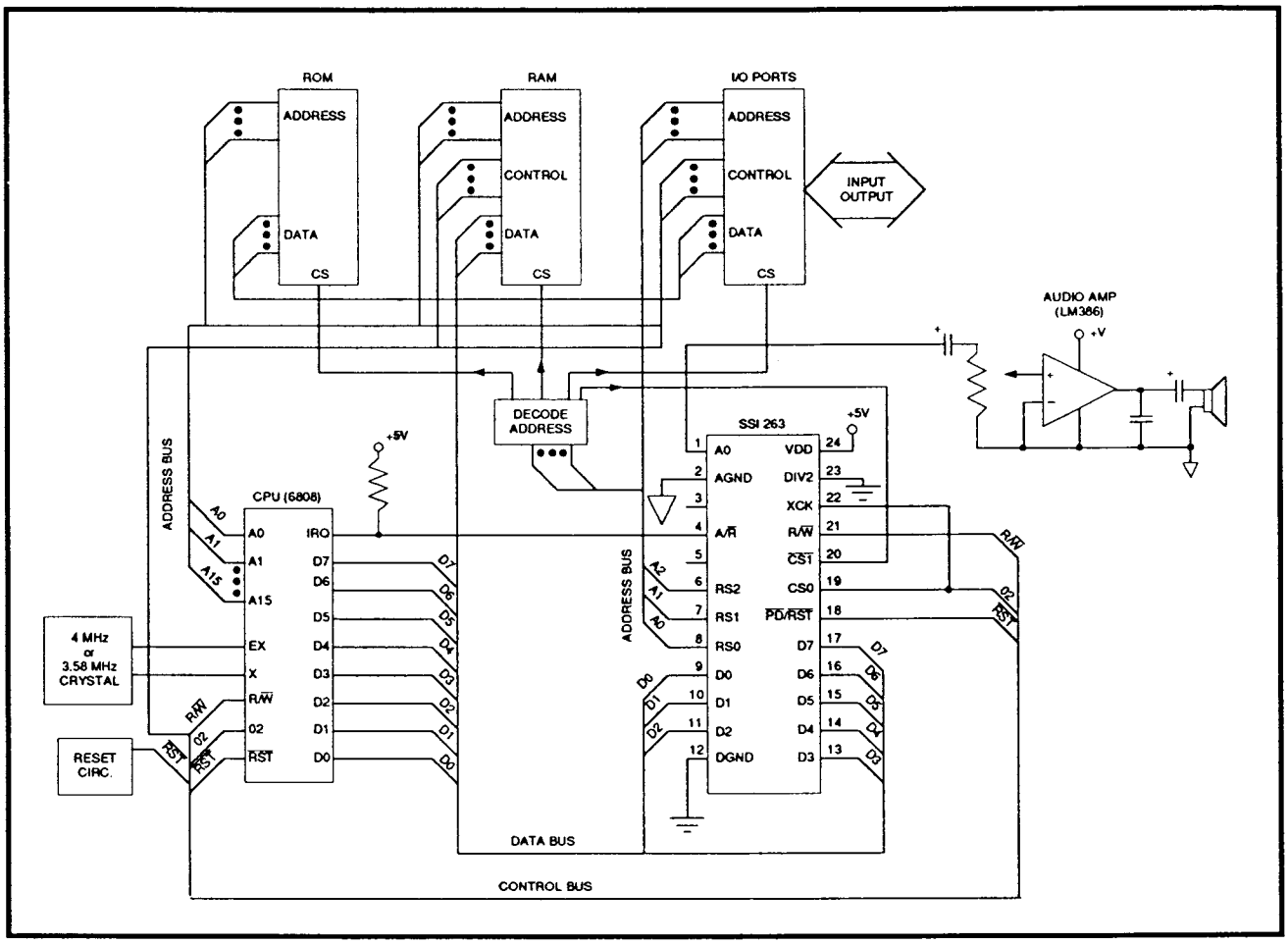
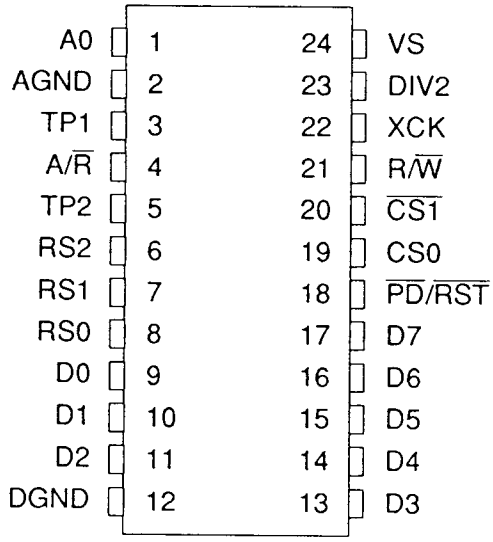


FIGURE 3: Typical Microprocessor Implementation

PACKAGE PIN DESIGNATIONS

(Top View)



24-Pin DIP

ORDERING INFORMATION

PART DESCRIPTION	ORDER NO.	PKG. MARK
SSI 78A263A 24-pin DIP	78A263A-CP	78A263A-CP

No responsibility is assumed by Silicon Systems for use of this product nor for any infringements of patents and trademarks or other rights of third parties resulting from its use. No license is granted under any patents, patent rights or trademarks of Silicon Systems. Silicon Systems reserves the right to make changes in specifications at any time without notice. Accordingly, the reader is cautioned to verify that the data sheet is current before placing orders.

Silicon Systems, Inc., 14351 Myford Road, Tustin, CA 92680 (714) 573-6000, FAX (714) 573-6914

©1993 Silicon Systems, Inc.