

- TOM GECKE - ?

silicon systems™

INNOVATORS IN INTEGRATION

SSI 263A Phoneme Speech Synthesizer

Data Sheet

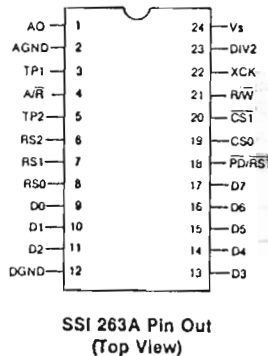
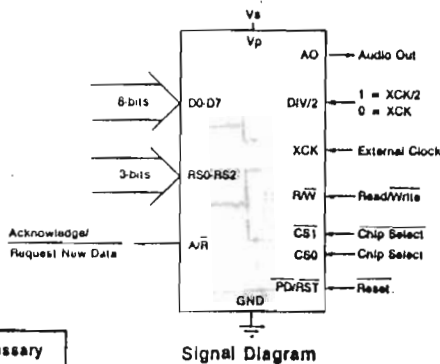
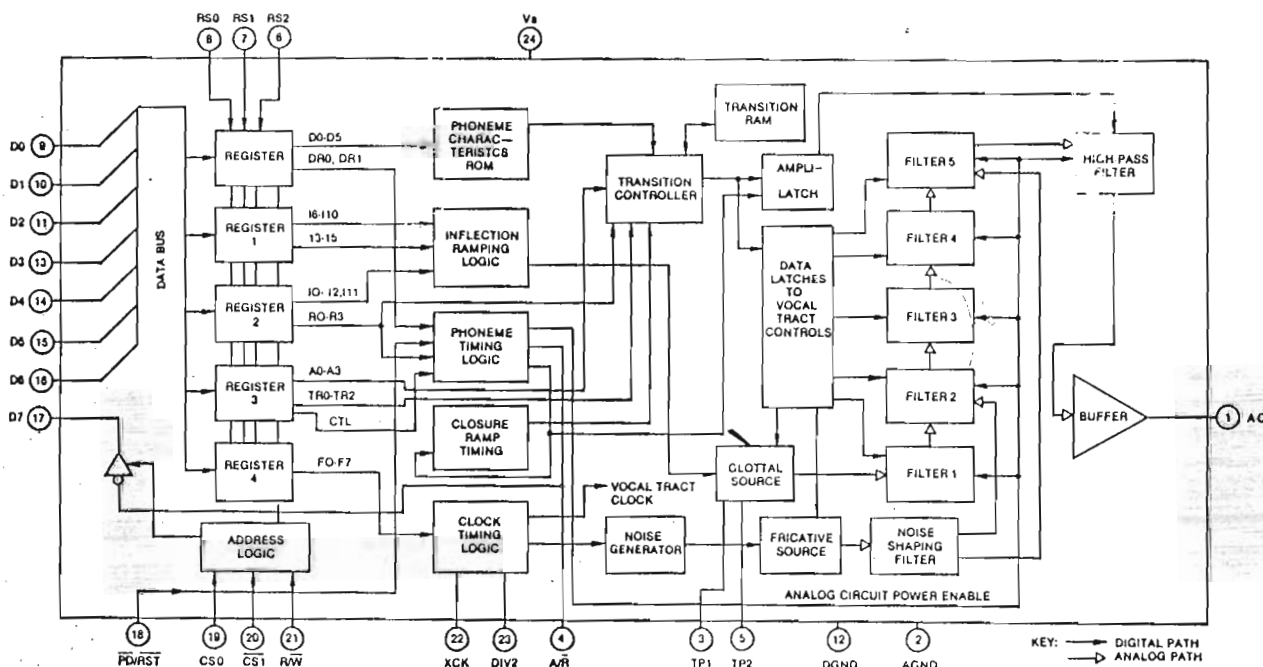
DESCRIPTION

The SSI 263A 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 263A 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 powerdown/reset mode
- Switched-capacitor-filter technology



CAUTION: Use handling procedures necessary for a static sensitive component

SSI 263A

SSI 263A Operation Description

This short description is intended to provide SSI 263A feature and capability information only. Refer to the SSI 263A USERS GUIDE for complete information on application and phonetic programming.

The Production of Speech

To produce different speech phonemes (sounds) the SSI 263A 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.

The 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 263A 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 263A 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, \overline{RW} (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 263A, 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 less than 100 bits-per-second. A higher data rate of

about 500 bits-per-second 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/R (Acknowledge/Request Not) pin active or disabled. Refer to the Mode Selection Chart.

Phoneme Selection

The SSI 263A 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 263A is in the mode of immediate inflection, bits I11-I0 provide immediate adjustment with seven octaves of pitch on an 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 90Hz where:

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

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 20KHz (see formula below), but may be manipulated to fine-tune speech quality or to change "voice type"; bass, baritone, etc.

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

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:

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

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:

$$\text{Phoneme Duration} = (\text{Frame Duration}) \times (4-D)$$

D = decimal equivalent of Duration Register setting
All internal attribute transitioning is performed relative to the Speech Rate Register setting. Speech rate does not effect 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 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 $\overline{\text{PD/RST}}$ 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 $\overline{\text{A/R}}$, when the device is put into Read ($\overline{\text{R/W}}$ is a logic 1 and the chip is selected, $\text{CS1} = 0$, $\text{CS0} = 1$). 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 specifications). It is desirable to establish a stable crystal controlled time base from 800 to 1000KHz 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 $\overline{\text{A/R}}$ 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.

PHONEME CHART

Hex Code*	Phoneme Symbol	Example Word (or Usage)
00	PA	(payee)
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

*Note — Hex codes shown with DR0, DR1 = 0 (longest Duration)

PIN ASSIGNMENT DESCRIPTIONS

Pin No.	Symbol	Active Level	Description
1	AO		Analog Audio Output biased @ VDD/2 requires an external audio amp for speaker drive
2	AGND		Analog Ground
3	TP1		Do not use
4	A/R		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 Pin 17 also.)
5	TP2		Do not use
6	RS2		Register Select Input — used to select one of five internal registers in conjunction with RS1 and RS0
7	RS1		Register Select (See pin 6)
8	RS0		Register Select (See pin 6)
9	D0		LSB of 8-bit data bus — input only
10	D1		Data Input (only)
11	D2		Data Input (only)
12	DGND		Digital Ground
13	D3		Data Input (only)

Pin No.	Symbol	Active Level	Description
14	D4		Data Input (only)
15	D5		Data Input (only)
16	D6		Data Input (only)
17	D7		MSB of 8-bit data bus. Bi-directional, inverse of pin 4 when read is high
18	PD/RST	Low	Power Down Control Input — Silences audio output and retains DC bias without disturbing register contents. Disables A/R output.
19	CS0	High	Chip Select Input
20	CS1	Low	Chip Select Input
21	R/W		Read/Write Control Input — Write is active low for loading internal registers. Read is active high but enables D7 only.
22	XCK		Clock Input (≈ 1 or 2 MHz)
23	DIV2	High	Clock Divide by Two — used when external clock is ≈ 2 MHz
24	VDD		Positive Voltage Supply

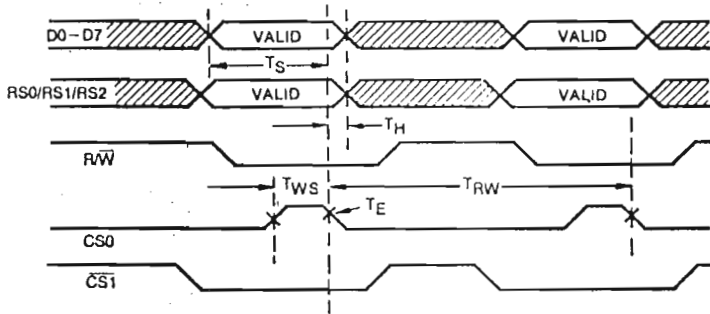
REGISTER INPUT FORMATS

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

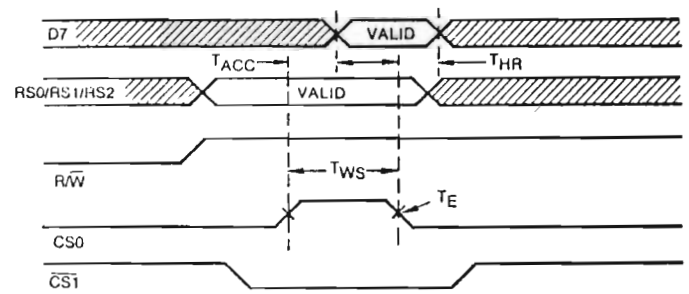
DR1, DR0 ... Define the phoneme duration.
 P5 → P0 ... Address the phoneme required.
 I11 → I0 ... Define inflection target frequencies and rate of change.
 R3 → R0 ... Define the rate or speed of speech.
 CTL ... Define the mode of A/R response in conjunction with DR1 and DR0.
 Also directly set by PD/RST.

T2 → T0 ... Define the rate of movement of the formant position for articulation purposes.
 A3 → A0 ... Define the amplitude of the output audio.
 F7 → F0 ... Define the frequency of all vocal tract filters.

WRITE TIMING DIAGRAM



READ TIMING DIAGRAM



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

Timing Characteristics

($V_{DD} = 4.5$ to 5.5 Volts, $T_A = -40$ to $+85$ deg. C)

Item	Symbol	Limits		Units.
		Min.	Max.	
Data Setup Time	TS	120**		nsec
Data Hold Time	TH	10**		nsec
Strobe Width	TWS	200		nsec
Read/Write Cycle Time	TRW	2.25*		μ sec
Rise/Fall Time	TE		100	nsec
D7 Output Access Time	TACC		180	nsec
D7 Output Hold Time	THR		180	nsec

Notes: * Based on color burst frequency.

** Timing relative to deselect by either $\overline{CS0}$, $\overline{CS1}$, or $\overline{R/W}$ changing.

MODE SELECTION CHART

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

ABSOLUTE MAXIMUM RATINGS

Item	Symbol	Limit	Units
Supply Voltage	$V_{DD}-V_{SS}$	7.0	V
Input Voltage	V_{IN}	-0.5 to $V_{DD} + 0.5$	V
D.C. Current at Inputs	I_{INM}	± 1.0	mA
Storage Temperature	T_S	-55 to $+125$	$^{\circ}$ C
Operating Temperature	T_A	-40 to $+85$	$^{\circ}$ C
Power Dissipation	P_d	500	mW

SSI 263A

Electrical Characteristics

Unless otherwise specified, $4.5 \leq V_{DD} \leq 5.5$; $-40 \text{ deg. C} \leq T_A \leq 85 \text{ deg. C}$;
 $1.50\text{MHz} \leq \text{XCK frequency} \leq 2.0\text{MHz}$, when $\text{XCK}/2 = \text{logic 1}$ or
 $0.75\text{MHz} \leq \text{XCK frequency} \leq 1.0\text{MHz}$, when $\text{XCK}/2 = \text{logic 0}$

Description	Conditions	Min.	Typ.	Max.	Units
POWER SUPPLY					
Supply Current	$\overline{\text{PD/RST}} = 1, \text{CTL} = 0$		8	20	mA
Supply Current	$\overline{\text{PD/RST}} = 0, \text{CTL} = 1$		7	18	mA
AUDIO OUTPUT					
Output Level	AW phoneme RL = 50Kohm to GND through $1\mu\text{F}$ cap.	0.28VDD	0.37VDD	0.50VDD	Vpp
DC Output Offset		0.5VDD	0.6VDD	0.7VDD	V
Resistive Loading	AC coupled to AO to GND	10			Kohm
Capacitive Loading	To GND to ensure Stable A			100	pF

Description	Conditions	Symbol	Min	Typ	Max	Units
BUS CONTROL INPUTS, DATA INPUTS (RS0, RS1, RS2, CS0, CS1, D0-D7 PD/RST)						
Input High Voltage		V_{IH}	$V_{SS} + 2.4$		$V_{DD} + 0.3$	VDC
Input Low Voltage		V_{IL}	-0.3		+0.8	VDC
Input Leakage Current	$V_{IN} = 0 \text{ to } V_{DD}$	I_{IN}			5	μA
Input Capacitance	$V_{IN} = 0$ $T_A = 25^\circ\text{C}$ measured at $f = 1.0\text{MHz}$	C_{IN}			10	pF
Input Capacitance, D7 Input		$C_{IN}(D7)$			20	pF
Input Current, D7 In TRI-State "OFF" State	$V_{IN} = 0.4 \text{ to } 2.4 \text{ V}$	$I_{IN}(TS)$		2.0	5.0	μA

D7 OUTPUT

D7 Output Low Voltage	$I_{Load} = 0.4 \text{ mA into D7}$	$V_{OL}(D7)$			0.4	VDC
D7 Output High Voltage	$I_{Load} = 205 \mu\text{A out of D7}$	$V_{OH}(D7)$		$V_{DD} - 2.0$		VDC

A/R OUTPUT

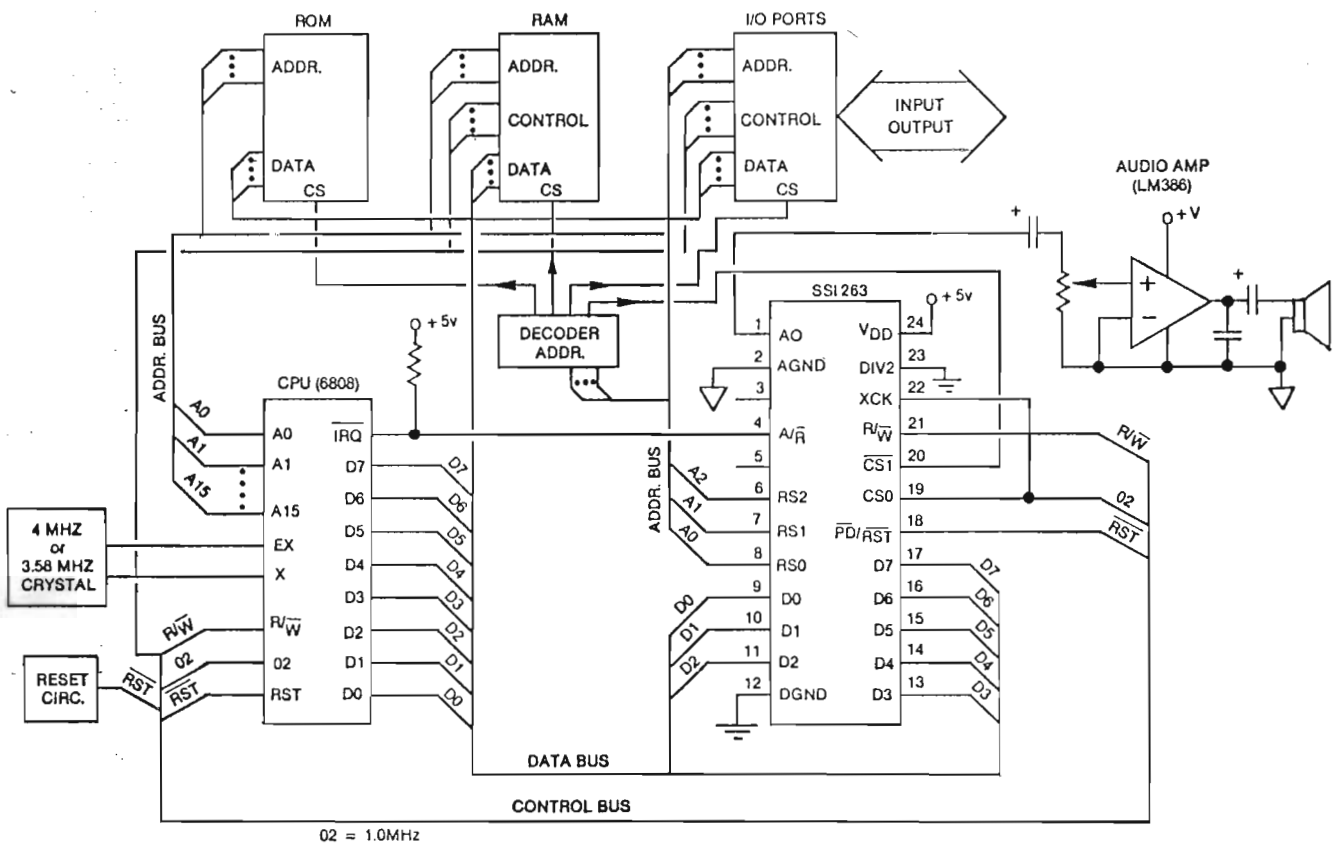
Output Low Voltage	$I_L = 3.2 \text{ mA into A/R}$	$I_{OL}(A/R)$			0.4	VDC
Output High Leakage Current	$V_{Out} = 0.0 \text{ to } V_{DD}$	$I_L(A/R)$			10	μA
Output Capacitance	$V_{Out} = 0 \text{ VDC}$ $T_{AMB} = 25^\circ\text{C}$ $f = 1.0 \text{ MHz}$	$C_{Out}(A/R)$		15	pF	

DIV2 INPUT

Input Low Voltage		$V_{IL}(DIV2)$	-0.3		.2 VDD	V
Input High Voltage		$V_{IH}(DIV2)$.8VDD		$V_{DD} + 0.3$	V
Input Leakage	$V_{IN} = 0 \text{ to } V_{DD}$				5	μA

Description	Conditions	Symbol	Min.	Typ.	Max.	Units.
XCLK						
Input Low Voltage		$V_{IH}(IC)$	-0.3		+0.8	V
Input High Voltage		$V_{IH}(IC)$	2.4		$V_{DD} + 0.3$	V
Input Current	$V_{IN} = 0.0$ to V_{DD}	$I_{IN}(C)$			5	μA
Input Capacitance		$C_{IN}(C)$			10	pF
Duty Cycle		$D(XCLK)$	0.4		0.6	—

TYPICAL MICROPROCESSOR IMPLEMENTATION





14351 Myford Road, Tustin, CA 92680 (714) 731-7110, TWX 910-595-2809

SSI STANDARD PRODUCTS TELECOMMUNICATIONS CIRCUITS

Part No.	Circuit Function	Characteristics	Voltage	Package
----------	------------------	-----------------	---------	---------

Tone Signaling Products

SSI 201	Integrated DTMF Receiver	Hexadecimal or binary 2-of-8 output	12V	22DIP
SSI 202	Integrated DTMF Receiver	Low power, hex or binary output	5 V	18 DIP
SSI 203	Integrated DTMF Receiver	Hex or binary output, Early Detect	5 V	18 DIP
SSI 204	Integrated DTMF Receiver	Low-power, binary output	5 V	14 DIP
SSI 957	Integrated DTMF Receiver	Early Detect, Dial Tone reject	5 V	22 DIP
SSI 20C89	Integrated DTMF Transceiver	Generator and Receiver, μ P interface	5 V	22 DIP
SSI 20C90	Integrated DTMF Transceiver	Generator and Receiver, μ P interface, Call Progress Detect	5 V	22 DIP
SSI 980	Call Progress Detector	Detects supervision tones, Teltone second-source	5 V	8 DIP

Modem Products

SSI K212	1200/300 Baud Modem	DPSK/FSK, single chip, autodial, Bell 212A	10V	28 DIP
SSI 223	1200 Baud Modem	FSK, HDX/FDX	10V	16 DIP
SSI 291/213	1200 Baud Modem	DPSK, two chips, low-pwer	10V	40/16 DIP
SSI 3522	1200 Baud Modem Filter	Bell 212 compatible, AMI second-source	10V	16 DIP

Speech Synthesis Products

SSI 263A	Speech Synthesizer	Phoneme-based, low data rate, VOTRAX second-source	5 V	24 DIP
----------	--------------------	--	-----	--------

Switching Products

SSI 80C50	T1 Transmitter	Bell D2, D3, D4, serial format and mux, low power	5 V	28 DIP, Q
SSI 80C60	T1 Receiver	Bell D2, D3, serial synchron, and demux, low power	5 V	28 DIP, Q
SSI 22100	Cross-point Switch	4x4x1, control memory, RCA second-source	12V	16 DIP
SSI 22101/2	Cross-point Switch	4x4x2, control memory, RCA second-source	12V	24 DIP
SSI 22106	Cross-point Switch	8x8x1, control memory, RCA second-source	5 V	28 DIP
SSI 22301	PCM Line Repeater	T1 carrier signal recondition	5 V	18 DIP

No responsibility is assumed by SSI 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 SSI. SSI reserves the right to make changes in specifications at any time and without notice.

Data Sheet

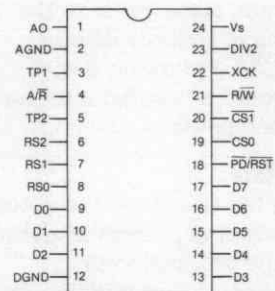
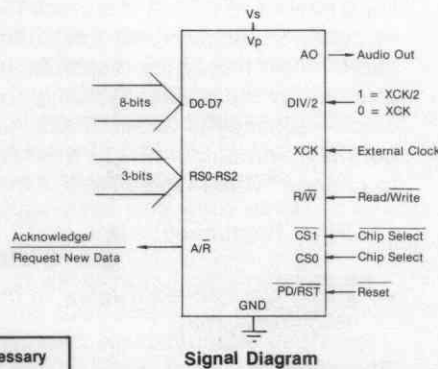
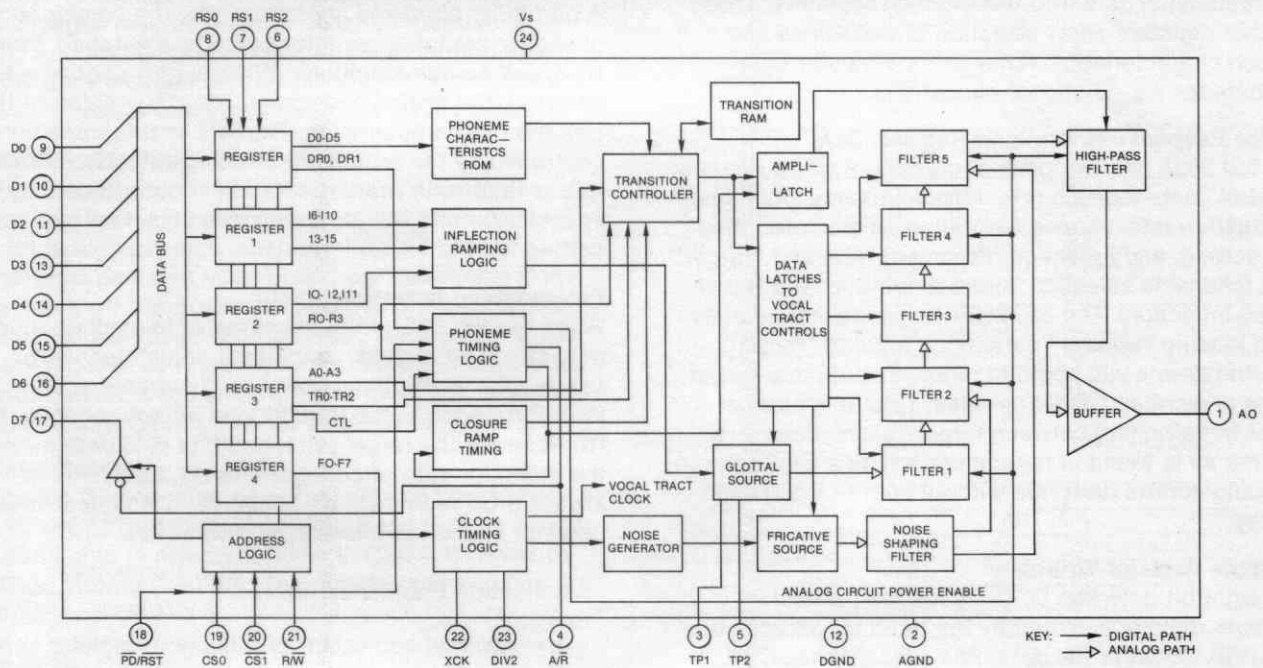
DESCRIPTION

The SSI 263A 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 263A 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 powerdown/reset mode
- Switched-capacitor-filter technology



CAUTION: Use handling procedures necessary for a static sensitive component

Signal Diagram

SSI 263A Pin Out (Top View)

SSI 263A

SSI 263A Operation Description

This short description is intended to provide SSI 263A feature and capability information only. Refer to the SSI 263A USERS GUIDE for complete information on application and phonetic programming.

The Production of Speech

To produce different speech phonemes (sounds) the SSI 263A 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.

The 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 263A 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 263A 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/\bar{W} (Read/Write), CS0 (Chip Select 0), and $\bar{CS}1$ 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 $\bar{CS}1$. 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 263A, 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 less than 100 bits-per-second. A higher data rate of

about 500 bits-per-second 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/ \bar{R} (Acknowledge/Request Not) pin active or disabled. Refer to the Mode Selection Chart.

Phoneme Selection

The SSI 263A 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 263A is in the mode of immediate inflection, bits I11-I0 provide immediate adjustment with seven octaves of pitch on an 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 90Hz where:

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

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 20KHz (see formula below), but may be manipulated to fine-tune speech quality or to change "voice type"; bass, baritone, etc.

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

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:

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

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:

$$\text{Phoneme Duration} = (\text{Frame Duration}) \times (4-D)$$

D = decimal equivalent of Duration Register setting

All internal attribute transitioning is performed relative to the Speech Rate Register setting. Speech rate does not effect 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 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 PD/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/R, when the device is put into Read (R/W is a logic 1 and the chip is selected, 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 specifications). It is desirable to establish a stable crystal controlled time base from 800 to 1000KHz 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.

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

*Note — Hex codes shown with DR0, DR1 = 0 (longest Duration)

SSI 263A

PIN ASSIGNMENT DESCRIPTIONS

Pin No.	Symbol	Active Level	Description
1	AO		Analog Audio Output biased @ $V_{DD}/2$ requires an external audio amp for speaker drive
2	AGND		Analog Ground
3	TP1		Do not use
4	A/R		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 Pin 17 also.)
5	TP2		Do not use
6	RS2		Register Select Input — used to select one of five internal registers in conjunction with RS1 and RS0
7	RS1		Register Select (See pin 6)
8	RS0		Register Select (See pin 6)
9	D0		LSB of 8-bit data bus — input only
10	D1		Data Input (only)
11	D2		Data Input (only)
12	DGND		Digital Ground
13	D3		Data Input (only)

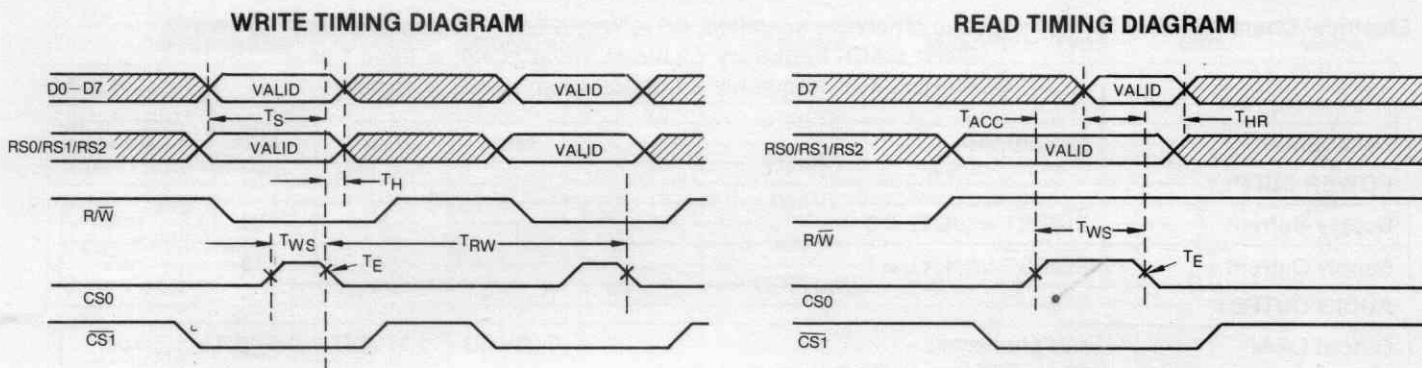
Pin No.	Symbol	Active Level	Description
14	D4		Data Input (only)
15	D5		Data Input (only)
16	D6		Data Input (only)
17	D7		MSB of 8-bit data bus. Bi-directional, inverse of pin 4 when read is high
18	PD/RST	Low	Power Down Control Input — Silences audio output and retains DC bias without disturbing register contents. Disables A/R output.
19	CS0	High	Chip Select Input
20	CS1	Low	Chip Select Input
21	R/W		Read/Write Control Input — Write is active low for loading internal registers. Read is active high but enables D7 only.
22	XCK		Clock Input (≈ 1 or 2 MHz)
23	DIV2	High	Clock Divide by Two — used when external clock is ≈ 2 MHz
24	VDD		Positive Voltage Supply

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/R response in conjunction with DR1 and DR0.
 Also directly set by PD/RST.

T2 → T0 ... Define the rate of movement of the formant position for articulation purposes.
 A3 → A0 ... Define the amplitude of the output audio.
 F7 → F0 ... Define the frequency of all vocal tract filters.



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

Timing Characteristics ($V_{DD} = 4.5$ to 5.5 Volts, $T_A = -40$ to $+85$ deg. C)

Item	Symbol	Limits		Units.
		Min.	Max.	
Data Setup Time	TS	120**		nsec
Data Hold Time	TH	10**		nsec
Strobe Width	TWS	200		nsec
Read/Write Cycle Time	TRW	2.25*		μ sec
Rise/Fall Time	TE		100	nsec
D7 Output Access Time	TACC		180	nsec
D7 Output Hold Time	THR		180	nsec

Notes: * Based on color burst frequency.

** Timing relative to deselect by either $\overline{CS0}$, $\overline{CS1}$, or $\overline{R/W}$ changing.

MODE SELECTION CHART

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

ABSOLUTE MAXIMUM RATINGS

Item	Symbol	Limit	Units
Supply Voltage	$V_{DD}-V_{SS}$	7.0	V
Input Voltage	V_{IN}	-0.5 to $V_{DD} + 0.5$	V
D.C. Current at Inputs	I_{INM}	± 1.0	mA
Storage Temperature	T_S	-55 to $+125$	$^{\circ}$ C
Operating Temperature	T_A	-40 to $+85$	$^{\circ}$ C
Power Dissipation	P_d	500	mW

SSI 263A

Electrical Characteristics

Unless otherwise specified, $4.5 \leq V_{DD} \leq 5.5$; $-40 \text{ deg. C} \leq T_A \leq 85 \text{ deg. C}$;
 $1.50\text{MHz} \leq \text{XCK frequency} \leq 2.0\text{MHz}$, when XCK/2 = logic 1 or
 $0.75\text{MHz} \leq \text{XCK frequency} \leq 1.0\text{MHz}$, when XCK/2 = logic 0

Description	Conditions	Min.	Typ.	Max.	Units
POWER SUPPLY					
Supply Current	$\overline{\text{PD}}/\overline{\text{RST}} = 1$, CTL = 0		8	20	mA
Supply Current	$\overline{\text{PD}}/\overline{\text{RST}} = 0$, CTL = 1		7	18	mA
AUDIO OUTPUT					
Output Level	AW phoneme RL = 50Kohm to GND through $1\mu\text{F}$ cap.	0.28VDD	0.37VDD	0.50VDD	Vpp
DC Output Offset		0.5VDD	0.6VDD	0.7VDD	V
Resistive Loading	AC coupled to AO to GND	10			Kohm
Capacitive Loading	To GND to ensure Stable A			100	pF

Description	Conditions	Symbol	Min	Typ	Max	Units
BUS CONTROL INPUTS, DATA INPUTS (RS0, RS1, RS2, CS0, CS1, D0-D7 PD/RST)						
Input High Voltage		V_{IH}	$V_{SS} + 2.4$		$V_{DD} + 0.3$	VDC
Input Low Voltage		V_{IL}	-0.3		+0.8	VDC
Input Leakage Current	$V_{IN} = 0$ to V_{DD}	I_{IN}			5	μA
Input Capacitance	$V_{IN} = 0$ $T_A = 25^\circ\text{C}$ measured at $f = 1.0\text{MHz}$	C_{IN}			10	pF
Input Capacitance, D7 Input		$C_{IN}(D7)$			20	pF
Input Current, D7 in TRI-State "OFF" State	$V_{IN} = 0.4$ to 2.4 V	$I_{IN}(TS)$		2.0	5.0	μA

D7 OUTPUT

D7 Output Low Voltage	$I_{Load} = 0.4$ mA into D7	$V_{OL}(D7)$			0.4	VDC
D7 Output High Voltage	$I_{Load} = 205$ μA out of D7	$V_{OH}(D7)$		$V_{DD} - 2.0$		VDC

A/R OUTPUT

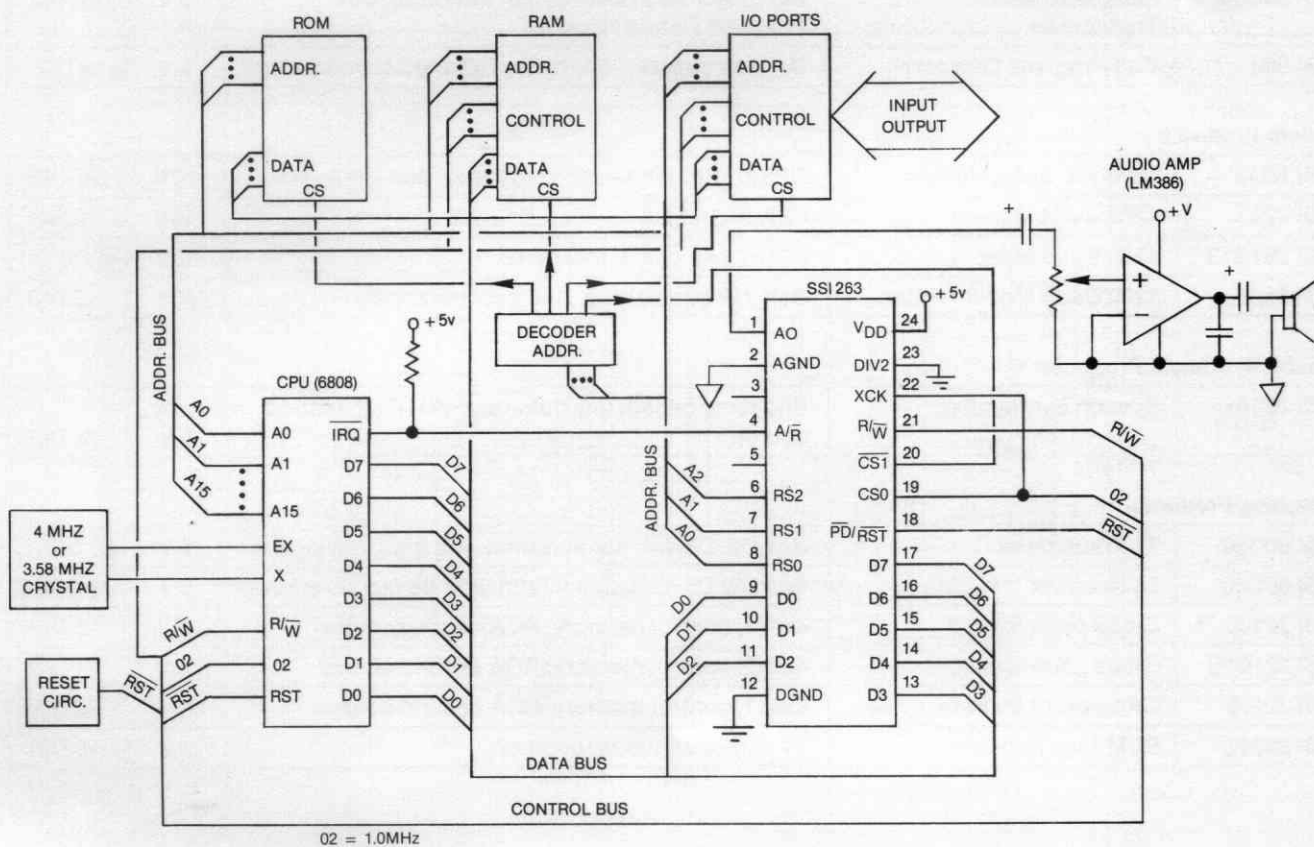
Output Low Voltage	$I_L = 3.2$ mA into A/R	$I_{OL}(A/\overline{R})$			0.4	VDC
Output High Leakage Current	$V_{Out} = 0.0$ to V_{DD}	$I_L(A/\overline{R})$			10	μA
Output Capacitance	$V_{Out} = 0$ VDC $T_{AMB} = 25^\circ\text{C}$ $f = 1.0$ MHz	$C_{Out}(A/\overline{R})$		15		pF

DIV2 INPUT

Input Low Voltage		$V_{IL}(DIV2)$	-0.3		.2 V_{DD}	V
Input High Voltage		$V_{IH}(DIV2)$.8 V_{DD}		$V_{DD} + 0.3$	V
Input Leakage	$V_{IN} = 0$ to V_{DD}				5	μA

Description	Conditions	Symbol	Min.	Typ.	Max.	Units.
XCLK						
Input Low Voltage		$V_{IH}(IC)$	-0.3		+0.8	V
Input High Voltage		$V_{IH}(IC)$	2.4		$V_{DD} + 0.3$	V
Input Current	$V_{IN} = 0.0$ to V_{DD}	$I_{IN}(C)$			5	μA
Input Capacitance		$C_{IN}(C)$			10	pF
Duty Cycle		$D(XCLK)$	0.4		0.6	—

TYPICAL MICROPROCESSOR IMPLEMENTATION



User's Guide for Phonetic Programming Using the SSI 263A

Phonetics

Every speech sound (phoneme) in any language may be represented by a special symbol (phonetic symbol). These symbols are used in WRITING precisely the sound sequence (phonetic transcription) of a word according to the way it is pronounced. There are many different phonetic symbol sets (phonetic alphabets). Each would contain a minimum number of symbols to represent the basic sounds (phonemes) required to pronounce any word in the language. Additional symbols are usually included which represent sounds with slight to great variations in the basic sounds (allophones). These symbols are used to assist in the transcription of words that reflect a regional, dialectic, or foreign pronunciation.

The process of transcribing a spoken word into its phonetic components begins with identifying the number of sounds in the word, then tagging each with a label to specify its type. Consonants and vowels are the most familiar labels but these may be broken down into subtypes (e.g., stop consonants, back vowels, etc.) as the need for more specificity arises. Once the sounds have been identified, their symbols are selected, then written in sequence. The resulting transcription should allow another person to identify the pronunciation without having heard the word spoken.

Note that when using a phonetic alphabet to transcribe words into their sound sequences, there is not a one-to-one correspondence between the alphabet characters (orthographics) used to spell words and the phonetic symbols (phonetics) used to represent their pronunciations. For example, in the word "phones" there are 6 letters but only 4 sounds. Conversely, the word "I" has 1 letter but 2 sounds. It may be of some assistance to keep a dictionary handy for reference. Dictionaries use their own phonetic system to describe the pronunciations of every word entry. It will be necessary to learn at least one phonetic alphabet in order to engage in phonetic transcription. The SSI 263A Phonetic Alphabet is the referent used in this manual. However, if another system is already known, it is easily translated into the referent.

When transcribing vocabulary from orthography (standard alphabet spelling) to phonetics, it is common to place the phonetic sequence between right slash marks when the transcription appears in running text. The word "phones," for example, would be transcribed as /F O N Z/ when using SSI 263A phonetic symbols. This allows the reader easier identification of phonetic segments.

SSI 263A Phonetic Alphabet

The phonetic alphabet used to represent the SSI 263A phonemes is the SSI 263A PHONETIC ALPHABET. Refer to the Phoneme Chart for a complete listing of the phoneme symbols.

Of the 64 alphanumeric symbols in the SSI 263A Phonetic Alphabet, 34 represent sound BASIC to the pronunciation of American English. The remaining 30 symbols fall into 2 groups: the ALLOPHONE group and the NO-SOUND group. The BASIC sound symbols are:

A, AE, AH, AW, B, D, E, EH, ER, F, HF, I, J, K, KV, L, M, N, NG, O, OO, P, R, S, SCH, T, TH, THV, U, UH, V, W, Y, Z.

Symbols in the ALLOPHONE group represent speech sounds that vary in pronunciation from one of the basic sounds. They may be used in transcribing words or word segments (syllables or morphemes) whose pronunciations are not satisfied by the basic phonemes alone (words rooted in a foreign language, words adapted by a regional dialect, etc.). The ALLOPHONE symbols are:

A1, AE1, AH1, AY, E1, E2, EH1, HN, HV, IE, IU, IU1, L1, LB, LF, OU, R1, R2, U1, UH1, UH2, UH3, Y1, :A, :OH, :U, :UH.

The NO-SOUND symbols represent silent states. One of these symbols represents a "pause" state. It is used to separate phoneme sequences into phrase-like segments which assist in more closely imitating the natural pausing in human speech for breathing or for delayed emphasis. The "pause" is treated as a phoneme when it is selected for a transcription and will be subject to phoneme parameter programming. It has the ability to maintain the parametric levels of duration, inflection, amplitude, etc., during its silence, thus audibly affecting the movement of the preceding and following phonemes. Other NO-SOUND symbols represent "hold" states. They are used in combination with BASIC phonemes or ALLOPHONES to generate articulation variations on their pronunciations. The NO-SOUND symbols are:

HFC, HVC, PA.

Now that there is a tool to use for writing the sounds that are heard, the next stage is to identify the sounds that are produced by the SSI 263A speech synthesizer.

SSI 263A Phoneme Review

Thus far in this program, it has been established that: (1) spoken words are made up of a series of sounds; (2) each speech sound in a language may be represented by an identifying symbol; and (3) the spoken word may be written according to its sound sequence using these special symbols. Before a word may be written phonetically, however, users may wish to study further the SSI 263A speech sounds. What makes one sound different from another and how these differences may be helpful to phonetic programming will be essential information for phonetic programmers.

The sound that is represented by each phonetic symbol in the SSI 263A Phonetic Alphabet must be audibly learned. The easiest way to approach this task is to start with the sounds already known and associate a symbol with them. For example, from spelling we have already learned that vowels may be "long" or "short" and are often differentiated by their particular spelling formats. Every time a word with a "short a" sound is heard (sat, fat, cat, bat, happy, plaster, ankle, Saturday, amplify, contaminate, etc.) the symbol /AE/ should come to mind. A "long a" sound (fate, state, bait, lace, maybe, stable, arrangement, etc.) is actually a diphthong (two sounds combined into a single unit) and may be represented by the symbols /A AY/.

In standard orthography, there are only 5 vowel letters to represent 17 vowel sounds. In phonetics, each vowel sound will be represented by its own symbol or symbol combination.

Again, from spelling, we have learned that the letter "c" may have a hard sound as in "cat" or a soft sound as in "city." The hard sound is actually a /K/ as in "kite" and the soft sound is an /S/ as in "sing." Users must identify which sound (/K/ or /S/) is used in the transcription of a "c." You will not find a symbol C in a phonetic alphabet. Like "C," the letters "Q" and "X" will not be found in phonetic alphabets. They are transcribed into the sound sequences /K W/ and /K PA S/. Refer to the Phoneme Chart during this learning period. It provides example words to describe the pronunciations corresponding to each symbol.

Users may add more words to the examples above to continue identifying the symbol-sound relationship for /AE/ and /A AY/. Follow this technique for each symbol in the alphabet. For auditory verification, enter the sound that is being reviewed into the device. Speak aloud your example word for the SSI 263A

sound in an attempt to match that which the synthesizer is emitting.

Example: /E/ = "long e" vowel sound
 = meat, read, need, repair, before, phoneme,
 erase, brief, people, timeliness, seniority,
 receive, catastrophe.

Example: /F/ = "voiceless fricative" consonant
 = farm, false, aft, feet, finger, phrase, phone,
 Africa, alphabet, cough.

Once you have reviewed auditorily the sounds you already have a familiarity with from spelling, proceed to the BASIC sound list in the above text and continue the review. Be aware that several consonant sounds will not provide output unless they have another sound following. This is the case with /B/, /D/, /P/, /T/, and /K/. When one of these sounds is entered into the SSI 263A, follow it by a vowel and listen to both in sequence.

Users who already have a familiarity with phonetics and SSI 263A synthetic sounds, may wish to follow the sound review procedures in order to auditorily determine the difference between two sounds or identify new ones. For example, enter the /UH/ phoneme into the device. Then enter /UH1/, /UH2/, and /UH3/. Listen to each sound noting the pronunciation variations. Be aware that there are no duplicate sounds resident on the SSI 263A chip.

Whenever a SSI 263A sound is audited that cannot be readily identified as to its appropriate usage, do not be concerned. The review is designed only to provide a method for establishing an auditory memory for each sound and a visual memory for its symbol. Phonetic programming may begin anytime after the initial review. Return to the review later as your familiarity with the BASIC sounds increases and as your need for sound alternatives to those BASIC sounds becomes more apparent.

If there is a question as to which symbols should be chosen to transcribe a word into its sound sequence, make a written note of the word by circling the letter(s) that present the problem. Later, when phonetic programming has begun, a phoneme sequence may be created for the word and users may verify auditorily which phonetic selection produces the most appropriate translation.

SSI 263A Phoneme Discussion

The SSI 263A Phonetic Alphabet is divided into 3 groups for the purpose of differentiating between phonemes and allophones. Another way of dividing the Alphabet is according to usage. The most familiar division is a two sections split: CONSONANT sounds and VOWEL sounds. Within each of these sections, sounds may be further subdivided according to the distinctive features that best describe the sounds phonetically or acoustically. The more that is known about a sound, the easier it is to determine how it may be used in transcribing and phonetically programming a word.

Consonant Sounds

There are 22 Consonant Phonemes, subdivided according to their manner of production in the human speech mechanism. Some are characterized by the noise emitted when the articulators obstruct the air flow (Fricatives like /S/). Vowel-like consonants have the least amount of obstruction and may occasionally be used as a vowel substitute. Stop consonants are obstructed completely, release of air flow occurring at the onset of the next sound. Notice that Affricates are a sequence of 2 sounds (a Stop followed by a Fricative) spoken as a single unit. Unlike vowels, which always have a vocal source during production, consonants may be voiced (V) or unvoiced (U) (no vocal source during air flow). When listening to the manner in which a consonant is produced during speech, note its special characteristics that distinguish it from all other consonants. The figure below displays all of the consonant sounds within their production groups.

	Stops	Fricatives	Affricates
Voiced	B, D, KV	Z, V, J, THV	(D, J)
Voiceless	P, T, K	S, F, SCH, TH, HF	(T, SCH)

	Semi-vowels	Glides	Nasals
Voiced	R, L	W, Y	M, N, NG
Voiceless			

Consonant Chart

Voiced and voiceless consonants are subdivided into 6 categories according to the manner in which they are produced in the human vocal tract: i.e., how the air flow is obstructed by the articulators to make each sound different.

Consonant sounds are selected for a sequence in much the same manner as an alphabet character would be selected for the spelling of a word. Users must be alert, however, to identify the exceptions. Occasionally, a consonant appears in the spelling of a word but not in its sound sequence: the "b" in "comb" is not pronounced and the sound sequence reflects the absence of the "b": /K OU M/. Some exceptions have grammatical rules that may be used in determining the appropriate sound. For example, a consonant may have 2 pronunciations according to its sound environment. The "s" used to pluralize the two words that follow are pronounced differently based on whether the sound that precedes it is voiced or unvoiced. An "s" pronunciation will match the voicing characteristics of the sound it follows.

Examples: tips = /T I P S/
 tabs = /T AE B Z/

There are other types of consonantal exceptions. For example, the "t" in a word like "nation" is pronounced /SH/ and the program might look like this: nation = /N A AY SH UH3 N/. Users must listen to each word's pronunciation to determine the appropriate phoneme selection.

There are 7 Consonant Allophones, each noted in the table below. The /L/ consonant is used in the initial position of a sequence for words beginning with "L", while the /LF/ allophone will occupy a medial or final position in a sequence: e.g., lull = /L UH LF/. The /LB/ and the /LI/ allophones would be used when a most constricted pronunciation of an "L" was required, as would occur for some words of foreign languages.

Consonant Phoneme	Consonant Allophones	Consonant Phoneme	Vowel Allophone
L	L1, LB, LF	R	ER
R	R1, R2	Y	YI

Allophone Listing for /L/, /R/, & /Y/

The /R/ is an initial position phoneme. Both /R1/ and /R2/ have more constricted pronunciations than /R/ and may be used in sequence with soundless interrupts to create a trilled /R/. Often when the /R/ is required in a medial or final position, it is vowelized and the /ER/ is used. Listening to the production of all four of these sounds will auditorily show that they may, occasionally, be used interchangeably.

Examples: red = /R EH D/
 bird = /B ER D/
 motor = /M OU T ER/

The /Y/ consonant, used as the final sound in words ending with "y," has a vowel allophone that may be used as the initial sound of words starting with "y." Note that both /Y/ and /YI/ are auditorily very close to the /E/ and the /IE/ vowels and may be considered interchangeable.

Vowel Sounds

There are 12 BASIC Vowel Phonemes. Vowels are subdivided according to the manner in which they are produced. All vowels are voiced sounds but each has a different output based on the degree of obstruction created by the opening of the mouth and the tongue position. Lip positions, another obstructing articulator, may range from spread flat to round. While the lips are in any of these positions, the jaw may be simultaneously dropped from a closed to an open position.

	Front Vowels	Medial Vowels	Back Vowels
	Spread	—————▶	Rounded
Closed ↓ Open	E		U
	I		OO
	A	UH	O
	EH	(ER)	AW
	AE		AH

Vowel Quadrilateral

Vowels begin their production with the same voiced energy. Changes in the position of the tongue (front or back), the shape of the lips (from spread flat to rounded), and the position of the lower jaw (from closed to open) determine the final characteristics that allow listeners to distinguish between vowel sounds.

Refer to the SSI 263A Phoneme Chart for the pronunciation reference on each BASIC vowel sound. Utilize the sound review techniques on the previous pages to practice identifying the vowel sounds in words and associating them with their phonetic symbols.

The allophonic variations of vowels, 20 in number, are used in a phonetic program to enhance the pronunciation of a word. There are some cases where the allophone is required for articulate pronunciations. This is true for /AY/, /YI/ and /IU/, which are integral components in the phonetic sequences for the "long a" and the varied "long u."

Examples: same = /S A AY M/
you = /YI IU U/

The table below places each allophone into the vowel quadrilateral to demonstrate approximately how they might relate to the BASIC vowels. Users are in no way restricted to traditional phonetic transcriptions that use only the BASIC vowel phonemes. Be encouraged to experiment with allophones. Place them in different positions in a sequence to auditorily check how they effect the overall pronunciation of a word.

	Front Vowels	Medial Vowels	Back Vowels
	Spread	—————▶	Rounded
Closed ↓ Open	YI E1 IE		U1
	AY	E2	IU IU1
	A1	UH1	OU
	EH1	UH2	
	AE1	UH3	AH1

Allophone Placement in Vowel Quadrilateral

Vowel allophones are placed in the vowel quadrilateral according to their production features. The sounds they emit vary slightly from the BASIC vowels that occupy the same positions.

Four vowel allophones—/A/, /OH/, /U/, and /UH/ — are adapted pronunciations of four of the BASIC vowels. These sounds are most commonly used for phonetically programming a foreign word. They may also be used as transitory sounds to link phonemes with opposite production features such as a round, open vowel with a very constricted, narrow consonant.

There are five vowels that require two or more vowel sounds in sequence in order to achieve their pronunciations. These are generally referred to as diphthongs. Refer to the Diphthong Conversion Chart.

The vowel quadrilateral is a handy tool to use for selecting vowel phonemes for diphthongs and other multi-phoneme units. For example, the diphthong in the word "I" starts with an /AH/ and ends with an /E/. In order to move smoothly from the first sound to the second (transition), another vowel may be inserted between these two sounds in sequence. The most likely choice would be a vowel that falls somewhere between /AH/ and /E/ in the quadrilateral: e.g., /UH/, /EH/, /I/, etc. The sequence may look like

this: /AH EH E/ or /AH1 UH3 IE/ or /AH1 EH3 AY/. In their fullest durations, a three-sound sequence would over articulate the diphthong. Shortening the first and last sounds by 1 duration and the medial sound by 2 durations will produce a more acceptable pronunciation (see SSI 263A Phoneme Parameters).

SSI 263A Phoneme Parameters (Attributes)

To achieve an accurate pronunciation of a word produced by the SSI 263A synthesizer requires more than a selection of the appropriate phonemes. Like human speech sounds, synthesized sounds are further defined by the rate at which they are emitted (duration), the level of pitch at which they are emitted (inflection or frequency), and the intensity with which they are produced (amplitude). These are considered the three major speech parameters which give the overall production of a word its linguistic character, transforming simple speech into more complex language. Inflection, amplitude, and duration are only three of the parameters that users have control of during the programming process. The rate at which one sound moves into another (articulation) is also a controllable parameter. Other parameters are: the slope of the inflection (slope), the rate of each selected duration (rate), and the extended inflection frequencies (extension). Users may also select the base frequency at which speech may be produced (filter frequency). Refer to SSI 263A Phoneme Parameters, for the range of each and typical default values selected.

Every phoneme selected for a sequence must be accompanied by assignments for each of the eight parameters. As users become more aware of their need to create different language effects with their synthesized speech output, they will require the flexibility and choice that comes with programmable parameters. For example, with 4 selectable durations per phoneme, each actual pronunciation of each sound may be changed. Thus, every sound has four possible outputs increasing the users' choice from 64 phonemes and allophones to 256. Each of the 256 may be effected differently by each of the 32 possible inflection assignments. Add to these possibilities 16 variations in amplitude and 16 variations in rate. The possible combinations are not limitless, of course, but they are very great and users are encouraged to experiment with as many as possible.

Several of the parameters effect synthetic speech output as a whole. These are articulation, pitch extension, and filter frequency. Users may select a single level at which to set the filter frequency, for example, and maintain that level throughout the programming process.

Phonetic Programming Methodology

Due to the great variety of phonemes and parameter choices, as well as the different effects the parameter selections have on the speech sounds, a systematic approach to selecting the variables is advised. The approach described below is only one of several that might be used. It may be adjusted to accommodate the user's special programming style or to accommodate later implementation of automatic control techniques.

The first step is to transcribe the target word, phrase, etc., into its basic phonetic components. Next, enter these sounds into the SSI 263A and auditorily check the output. Use the default values suggested in the Nominal Phoneme Parameter Table. The results should be a bit stilted if not misarticulated for the first trial program. Phoneme adjustment is next. Continue to make changes in the phoneme sequence, auditorily monitoring the changes, until an adequate pronunciation of the target is established.

Begin parameter adjustments. First, maintain articulation, pitch extension and filter frequency at nominal values. The device should be kept in the transitioned inflection mode. Make adjustments in the levels of only one of the remaining 4 parameters at a time, beginning with the duration and moving on to the inflection, rate, and amplitude (in that order) once the specific effect that the parameter can make has been made. Return to a previously adjusted parameter at any time based on need.

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

SSI 263A Diphthong Conversion Chart

Phoneme Sequence	Example Words
A AY Y	rain, became, stay
A IE EH1 UH3 LF	mail, hale, avail
AH1 AE1 EH1 Y	time, rhyme, sky
AH1 EH1 IE AW UH3 LF	smile, style, while
AH1 EH1 IE UH3 ER	fire, liar, inspire
UH3 AH1 Y	mice, right, sniper
O U	road, stone, lower
OU O O	tore, four, floor
AH1 AW O U	loud, flower, hour
UH3 AH1 O U	house, about, ouch
O UH1 AH1 I IE	boy, noise, annoy
O UH3 EH1 I OO LF	boil, spoil, doily
IU U U	tune, spoon, do
YI IU U U	you, few, music

SSI 263A Multi-Unit Conversion Chart

Phoneme Sequence	Example Words
T HFC SCH	church, latch
KV HVC HF	good, lag, angry
D J	just, ledge, wage
KV HF HFC	lake, corn, check
P HF	pipe, pay, poor
K HF W	quest, quick, aqua
T HF	top, trip, strain
HFC K HF HVC S	six, exit, taxi

Nominal Phoneme Parameter Table (Suggested Default Values for Speech Development)

Amplitude (A3 → A0)

Range—0 to F (softest to loudest, 0 = silent)

Default—C

Exceptions—KV = 0, B = D = 6

Duration (DR1, DR0)

Range—3 to 0 (shortest to longest)

Default—0

Filter Frequency Range (F7 → F0)

Range—00 to FF (lowest to highest)

Default—E9

Inflection (Pitch) (I10 → I6, Transitioned Inflection Mode Only)

Range—0 to 1F (lowest to highest, 0 = silent)

Default—04

Extension and Range of Pitch (I11, I2 → I0)

Range—0 to 7 (low); 8 to F (high)

Default Value—8

Rate of Speech (R3 → R0)

Range—0 to F (slowest to fastest)

Default—A

Slope of Inflection (I6 → I3, Transitioned Inflection Mode Only)

Range—0 to 7

Default—0

Articulation (Rate of) (A3 → A0)

Range—0 to 7 (slow to fast)

Default—5

*Note — Hex codes shown with DR0, DR1 = 0 (longest Duration)

Example of Using Phonetic Programming Methodology:

Developing "Hello"

Phoneme Parameters					SSI 263 Register Data				
Pho.D	T	In-S	A	R E FF	DP	IS	RE	TA	FF

KEY: Pho = Phoneme
 D = Duration
 T = Articulation
 In = Inflection
 S = Slope of Inflection
 A = Amplitude
 R = Rate
 E = Extension and Range of Pitch
 FF = Filter Frequency

DP = Duration/Phoneme Register Address 000
 IS = Inflection/Slope Register 001
 RE = Rate/Extension Register 010
 TA = Articulation/Amplitude Register 011
 FF = Filter Frequency Register 1XX

1. Original Phoneme Entry:

Pho.D	T	In-S	A	R E FF	DP	IS	RE	TA	FF
PA	.0	5	0A-0	C A 8 E9	00	50	A8	5C	E9
PA	.0	5	0A-0	C A 8 E9	00	50	A8	5C	E9
HF	.0	5	0A-0	C A 8 E9	2C	50	A8	5C	E9
EH	.0	5	0A-0	C A 8 E9	0A	50	A8	5C	E9
L	.0	5	0A-0	C A 8 E9	20	50	A8	5C	E9
O	.0	5	0A-0	C A 8 E9	11	50	A8	5C	E9
PA	.0	5	0A-0	C A 8 E9	00	50	A8	5C	E9
PA	.0	5	0A-0	C A 8 E9	00	50	A8	5C	E9

2. Phoneme Selection Refinement

Pho.D	T	In-S	A	R E FF	DP	IS	RE	TA	FF
PA	.0	5	0A-0	C A 8 E9	00	50	A8	5C	E9
PA	.0	5	0A-0	C A 8 E9	00	50	A8	5C	E9
HF	.0	5	0A-0	C A 8 E9	2C	50	A8	5C	E9
EH	.0	5	0A-0	C A 8 E9	0A	50	A8	5C	E9
UH3	.0	5	0A-0	C A 8 E9	1B	50	A8	5C	E9
LF	.0	5	0A-0	C A 8 E9	22	50	A8	5C	E9
UH3	.0	5	0A-0	C A 8 E9	1B	50	A8	5C	E9
O	.0	5	0A-0	C A 8 E9	11	50	A8	5C	E9
OU	.0	5	0A-0	C A 8 E9	12	50	A8	5C	E9
U	.0	5	0A-0	C A 8 E9	16	50	A8	5C	E9
PA	.0	5	0A-0	C A 8 E9	00	50	A8	5C	E9
PA	.0	5	0A-0	C A 8 E9	00	50	A8	5C	E9

3. Duration Adjustment

Pho.D	T	In-S	A	R E FF	DP	IS	RE	TA	FF
PA	.0	5	0A-0	C A 8 E9	00	50	A8	5C	E9
PA	.0	5	0A-0	C A 8 E9	00	50	A8	5C	E9
HF	.1	5	0A-0	C A 8 E9	6C	50	A8	5C	E9
EH	.0	5	0A-0	C A 8 E9	0A	50	A8	5C	E9
UH3	.2	5	0A-0	C A 8 E9	9B	50	A8	5C	E9
LF	.0	5	0A-0	C A 8 E9	22	50	A8	5C	E9
UH3	.2	5	0A-0	C A 8 E9	9B	50	A8	5C	E9
O	.2	5	0A-0	C A 8 E9	91	50	A8	5C	E9
OU	.0	5	0A-0	C A 8 E9	12	50	A8	5C	E9
U	.3	5	0A-0	C A 8 E9	D6	50	A8	5C	E9
PA	.0	5	0A-0	C A 8 E9	00	50	A8	5C	E9
PA	.0	5	0A-0	C A 8 E9	00	50	A8	5C	E9

4. Phoneme and Duration Adjustment

Pho.D	T	In-S	A	R E FF	DP	IS	RE	TA	FF
PA	.0	5	0A-0	C A 8 E9	00	50	A8	5C	E9
PA	.0	5	0A-0	C A 8 E9	00	50	A8	5C	E9
HF	.1	5	0A-0	C A 8 E9	6C	50	A8	5C	E9
EH1	.1	5	0A-0	C A 8 E9	4B	50	A8	5C	E9
UH3	.2	5	0A-0	C A 8 E9	9B	50	A8	5C	E9
LF	.0	5	0A-0	C A 8 E9	22	50	A8	5C	E9
UH3	.2	5	0A-0	C A 8 E9	9B	50	A8	5C	E9
O	.2	5	0A-0	C A 8 E9	91	50	A8	5C	E9

OU	.0	5	0A-0	C A 8 E9	12	50	A8	5C	E9
U	.2	5	0A-0	C A 8 E9	96	50	A8	5C	E9
PA	.0	5	0A-0	C A 8 E9	00	50	A8	5C	E9
PA	.0	5	0A-0	C A 8 E9	00	50	A8	5C	E9

5. Inflection Adjustment

Pho.D	T	In-S	A	R E FF	DP	IS	RE	TA	FF
PA	.0	5	0B-0	C A 8 E9	00	58	A8	5C	E9
PA	.0	5	0B-0	C A 8 E9	00	58	A8	5C	E9
HF	.1	5	0A-0	C A 8 E9	6C	50	A8	5C	E9
EH1	.1	5	0B-0	C A 8 E9	4B	40	A8	5C	E9
UH3	.2	5	09-0	C A 8 E9	9B	48	A8	5C	E9
LF	.0	5	0B-0	C A 8 E9	22	40	A8	5C	E9
UH3	.2	5	05-0	C A 8 E9	9B	28	A8	5C	E9
O	.2	5	05-0	C A 8 E9	91	28	A8	5C	E9
OU	.0	5	06-0	C A 8 E9	12	30	A8	5C	E9
U	.2	5	07-0	C A 8 E9	96	38	A8	5C	E9
PA	.0	5	0A-0	C A 8 E9	00	50	A8	5C	E9
PA	.0	5	0B-0	C A 8 E9	00	58	A8	5C	E9

6. Phoneme, Duration, Inflection, and Rate Adjustment

Pho.D	T	In-S	A	R E FF	DP	IS	RE	TA	FF
PA	.0	5	0B-0	C A 8 E9	00	58	A8	5C	E9
PA	.0	5	0B-0	C A 8 E9	00	58	A8	5C	E9
HF	.1	5	0A-0	C 7 8 E9	6C	50	78	5C	E9
EH1	.1	5	0B-0	C D 8 E9	4B	40	D8	5C	E9
UH3	.2	5	09-0	C C 8 E9	9B	48	C8	5C	E9
LF	.0	5	0B-0	C C 8 E9	22	40	C8	5C	E9
UH3	.2	5	05-0	C 9 8 E9	9B	28	98	5C	E9
O	.2	5	05-0	C 9 8 E9	91	28	98	5C	E9
OU	.0	5	06-0	C A 8 E9	12	30	A8	5C	E9
U	.2	5	07-0	C C 8 E9	96	38	C8	5C	E9
U	.3	5	0A-0	C 7 8 E9	D6	50	78	5C	E9
PA	.0	5	0B-0	C A 8 E9	00	58	A8	5C	E9
PA	.0	5	0A-0	C A 8 E9	00	50	A8	5C	E9

7. Phoneme, Duration, Inflection, Rate, and Amplitude Adjustment

Pho.D	T	In-S	A	R E FF	DP	IS	RE	TA	FF
PA	.0	5	0B-0	C A 8 E9	00	58	A8	5C	E9
PA	.0	5	0B-0	C A 8 E9	00	58	A8	5C	E9
EH	.0	5	07-0	0 D 8 E9	0A	38	D8	50	E9
HF	.1	5	0A-0	4 7 8 E9	6C	50	78	54	E9
EH1	.1	5	0B-0	C D 8 E9	4B	40	D8	5C	E9
UH3	.2	5	09-0	A C 8 E9	9B	48	C8	5A	E9
LF	.0	5	0B-0	A C 8 E9	22	40	C8	5A	E9
UH3	.2	5	05-0	C 9 8 E9	9B	28	98	5C	E9
O	.2	5	05-0	C 9 8 E9	91	28	98	5C	E9
OU	.0	5	06-0	C A 8 E9	12	30	A8	5C	E9
U	.2	5	07-0	A C 8 E9	96	38	C8	5A	E9
U	.3	5	0A-0	0 7 8 E9	D6	50	78	50	E9
PA	.0	5	0B-0	C A 8 E9	00	58	A8	5C	E9
PA	.0	5	0A-0	C A 8 E9	00	50	A8	5C	E9

8. Further Adjustment (depending on personal preference)

Pho.D	T	In-S	A	R E FF	DP	IS	RE	TA	FF
PA	.0	5	0D-0	C A 8 E9	00	68	A8	5C	E9
PA	.0	5	0D-0	C A 8 E9	00	68	A8	5C	E9
EH	.0	5	0D-0	0 D 8 E9	0A	68	D8	50	E9
HF	.1	5	07-0	2 8 8 E9	6C	38	88	52	E9
EH1	.1	5	09-2	C D 8 E9	4B	4A	D8	5C	E9
UH3	.2	5	09-4	A C 8 E9	9B	4C	C8	5A	E9
LF	.0	5	09-0	A C 8 E9	22	48	C8	5A	E9
UH3	.2	5	07-7	C 9 8 E9	9B	3F	98	5C	E9
O	.2	5	06-4	C 9 8 E9	91	34	98	5C	E9
OU	.1	5	05-2	C A 8 E9	52	2A	A8	5C	E9
U	.2	5	06-3	3 5 8 E9	96	33	58	53	E9
U	.3	5	07-4	0 C 8 E9	D6	3C	C8	50	E9
PA	.0	5	05-4	C C 8 E9	00	2C	C8	5C	E9
PA	.0	5	01-4	C C 8 E9	00	0C	C8	5C	E9