

Voice Direct™ Speech Recognition IC

Speaker-Dependent Speech Recognition Solution

GENERAL DESCRIPTION

Voice Direct™, from the Interactive Speech™ family of products, is a speaker-dependent speech recognition IC designed for use in cost-sensitive electronic products. In addition to performing speech recognition, Voice Direct plays speech prompts, performs system control functions, provides status outputs and interfaces to external ROM and Serial EEPROM. Voice Direct can be controlled by an external host processor, or it can operate in a pin configurable stand-alone mode.

Voice Direct employs a sophisticated neural network to recognize trained words or phrases with greater than 99% accuracy. The highly-integrated nature of the chip reduces external parts count. A complete recognition system can be built with few additional parts other than a battery, speaker, external memory, microphone, crystal, and audio input circuitry. Voice Direct is available as an IC or in a complete system module that includes a PCB and all external components.

Voice Direct can be easily integrated into existing products or used to quickly design new products. High quality speech recognition is now possible in cost-sensitive consumer products!

FEATURES

High Quality, Low Cost Speech Recognition

- Speaker-dependent recognition works in any language
- Recognizes up to 60 words/phrases
- Supports phrases up to 3.2 seconds
- Minimal memory less than 100 bytes/word external

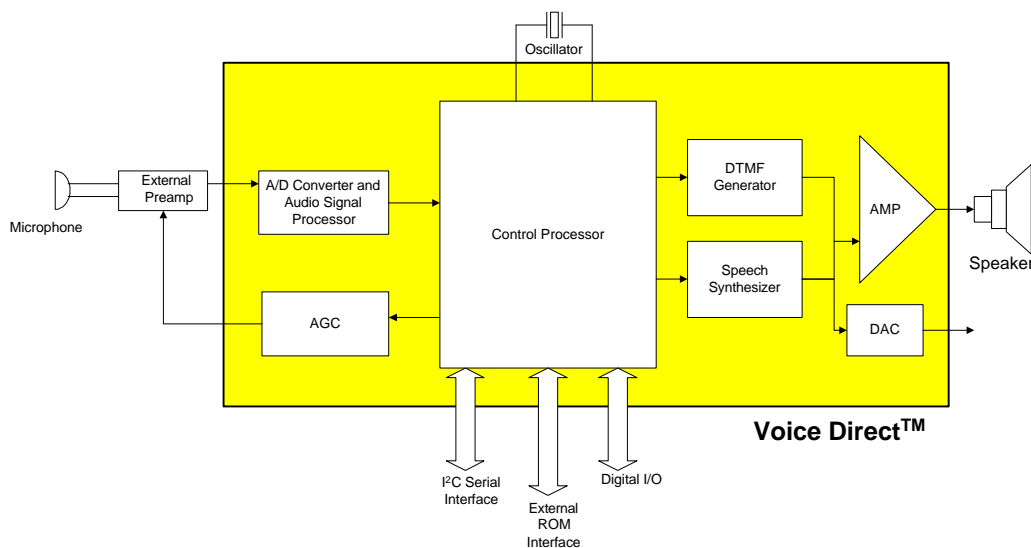
Integrated Single-Chip Solution

- Direct interface to 8K byte external memory for template storage (Serial EEPROM)
- English speech prompts
- On-chip A/D and D/A
- Output PWM circuitry for direct speaker drive
- Language localization and custom synthesis options

Two Operating Modes

- External Host - controlled by an external processor through a simple 3-wire host interface
- Stand Alone - pin-configurable operation

Voice Direct™ Block Diagram



FEATURE OVERVIEW

Voice Direct performs high quality speaker-dependent speech recognition. The chip utilizes its neural network recognizer to recognize discrete words or short phrases. The chip performs three basic functions:

Train - Users train the chip to identify a specific word by saying each word twice. After training, the two patterns are averaged and a template is stored.

Recognize - The user speaks a word and the chip compares the new pattern with the previously trained templates to identify which word was spoken. The chip then outputs the result of its analysis.

Erase - Users can delete previously trained words from the set of recognition templates.

In each of these functions, Voice Direct features integrated speech prompting providing a complete interactive user interface.

EXTERNAL HOST MODE

Voice Direct's external host operating mode provides a complete speaker dependent recognition system that can easily be controlled by an External Host processor (Host). The Host communicates to Voice Direct using a 3-wire serial bus. This high-level control interface allows the Host to control the flow of operations and to initiate all of its functions including training, recognition, or synthesis. In external host mode, Voice Direct recognizes up to 60 words. To improve application flexibility these words can be divided into smaller recognition sets, improving accuracy and functionality.

STAND ALONE MODE

Voice Direct's stand alone operating mode is designed to provide a complete recognition system using only the chip, external template storage memory, and a few passive electronic components. All operations, including training, recognition, and erase can be controlled by configuring chip input pins. Output pins provide status information to external devices. In stand alone mode, Voice Direct can recognize one set of 15 words.

SPEECH PROMPTS

Voice Direct includes a standard English vocabulary of over 100 phrases to guide the user through its functions. This standard word list can be replaced with a customized

word list for English or foreign languages via an external ROM chip.

RECOGNITION THRESHOLD

Voice Direct supports multiple acceptance threshold levels during the recognition process. The acceptance level determines how closely the spoken word must match a pre-trained template in order to pass. The user adjusts the level depending on the complexity of the recognition set. More complex recognition sets should have a higher acceptance level, while simpler sets can use a lower threshold level.

INPUT AUDIO AMPLIFIER AND FILTER

Voice Direct requires an external pre-amplifier to condition the input signal. When used with an inexpensive omni-directional electret microphone, the input audio amplifier and filter must provide approximately 58 dB of low-noise mid-band gain, 2-bit AGC controllability, and a first order bandpass response with 3dB points at roughly 700 Hz and 3300 Hz.

AUDIO OUTPUT

Voice Direct can directly drive a 32-Ohm speaker from the SP0 pin, providing approximately 0.15W of audio power.

MEMORY INTERFACES

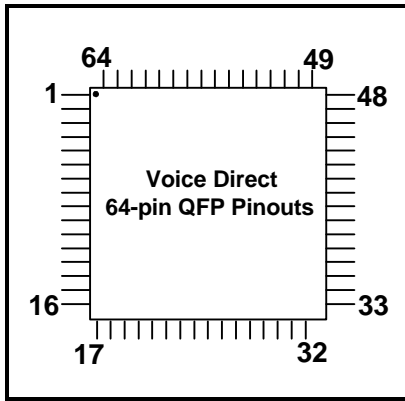
Voice Direct requires 8K bytes of dedicated external Serial EEPROM memory for template storage. Each time a new word is trained, Voice Direct automatically writes the template to the memory device. During recognition, Voice Direct reads the templates from the memory device and compares them with spoken words or phrases. Voice Direct communicates through a I²C 2-wire serial interface.

TSSP MODULE

The Voice Direct solution is also available as a complete module. The module is a single 2" x 2" PCB that includes all external components (e.g., preamplifier, memory) required by Voice Direct, except microphone and speaker. This module is ideal for prototype development or small production runs.

Feature Summary Of Voice Direct™

	Maximum Number of Recognition Words	Multiple Recognition Sets Supported	Acceptance Threshold Levels	Custom Synthesis	Foreign Language Synthesis
External Host	60	Yes (up to 8)	5	Yes	Yes
Stand-Alone	15	No	3	No	Yes



ELECTRICAL SPECIFICATIONS

OPERATING CONDITIONS -20°C to +70°C;
 $V_{cc}=3.5-5.0V$;
 $V_{ss}=0V$

ABSOLUTE MAXIMUM RATINGS

Maximum voltage 7.5V
 Minimum voltage on any pin $V_{ss}-0.6V$
 Maximum voltage on any pin $V_{dd}+0.6V$
 Any pin to GND ~0.1V to +7.5V
 Operating temperature (T_O) -20°C to +70°C
 Soldering temperature 260°C for 10 sec

WARNING: Stressing the Voice Direct beyond the “Absolute Maximum Ratings” may cause permanent damage. These are stress ratings only. Operation beyond the “Operating Conditions” is not recommended and extended exposure beyond the “Operating Conditions” may affect device reliability.

Name	Pin	Description	I/O
A[15:0]	1-8, 11-18	External ROM Memory Address Bus	O
AGC0	33	AGC control 0. The Voice Direct controls the amplifier gain with this signal.	O
AGC1	32	AGC control 1	O
AGND	52	Analog Ground. For noise reasons, analog and digital grounds should be separate.	-
AIN0	51	Analog In, low gain. (range AGND to AVDD/2.)	I
AIN1	50	Analog In, hi gain (8X input amplitude of AIN0, same range)	I
AVDD	55	Analog Voltage. For noise reasons, keep this supply independent of digital circuitry.	-
DACOUT	48	Analog Output (unbuffered).	O
GND	9, 22, 41, 56	Digital Ground	-
MD[7:0]	57-64	External ROM Memory Data Bus	I/O
MEM1	35	Memory Control 1. Serial Clock for Serial EEPROM.	O
MEM2	34	Memory Control 2. Serial Data for Serial EEPROM.	I/O
MODE/SP 1	54	The MODE pin is used to select Stand alone or CPU mode. This pin is also Speaker Connect1. A 32-Ohm speaker can be connected directly to this pin.	O (I at powerup)
-RESET	21	Reset	I
-RM	44	Read Memory Strobe. Can control -OE pin of External ROM.	O
SH	49	Sample and Hold. Connect a 470 pF capacitor from here to AGND.	I
SP0	53	Speaker Connect0. A 32-Ohm speaker can be connected directly to this pin.	O
VDD	10,23,36,40, 46, 47	Digital Supply Voltage (core)	-
-WR	43	Write Result. After a recognition sequence the chip places the result on the memory data bus MD[7:0] and strobes this signal to latch the result into external devices.	O
X1, X2	19,20	Crystal connect. A 14.312 mHz crystal is connected to these pins.	O
NC	42	*** NO CONNECT	-
Name	Pin	Description : I/O	
	64 MQFP	EXTERNAL HOST MODE	STAND ALONE MODE
HIGH	24	I/O: expansion bit 3	O: Add 8 to selected Category
OUT7	25	I/O: expansion bit 2	O: Category 7 (or 15 with HIGH)
OUT6	26	I/O: expansion bit 1	O: Category 6 (or 14 with HIGH)
OUT5	27	I/O: expansion bit 0	O: Category 5 (or 13 with HIGH)
OUT4	28	I/O: External host Bus Data bit3	O: Category 4 (or 12 with HIGH)
OUT3	29	I/O: External host Bus Data bit2	O: Category 3 (or 11 with HIGH)
OUT2	30	I/O: External host Bus Data bit1	O: Category 2 (or 10 with HIGH)
OUT1	31	I/O: External host Bus Data bit0	O: Category 1 (or 9 with HIGH)
ERROR	36	O: Voice Direct sets LOW to indicate processing in progress	O: Indicates an error occurred during the last training or recognition sequence.
-TRAIN	37	O: Auxiliary status bit.	I: “L”: Initiate Set Training.
-RECOG	38	I: CPU sets LOW to request action	I: “L”: Initiate Recognition
MHS	39	I: Master handshake pin	

THE INTERACTIVE SPEECH™ PRODUCT LINE

The Interactive Speech line of ICs and software was developed to “bring life to products” through advanced speech recognition and audio technology. The Interactive Speech Product Line was designed for consumer telephony products and cost-sensitive consumer electronic applications such as home electronics, personal security, and personal communication. The product line includes award-winning RSC-series general purpose microcontrollers plus a line of easy-to-implement chips which can be pin-configured or controlled by an external host microcontroller. Sensory’s software technologies run on a variety of microcontrollers and DSPs.

RSC-164

The RSC-164 is a low-cost 8-bit microcontroller designed for use in consumer electronics. It is a fully integrated microcontroller and includes A/D, D/A, ROM, and RAM circuitry on chip. The RSC-164 can perform a full range of speech/audio functions including speech recognition, speaker verification, speech and music synthesis, and voice record/playback.



Voice Direct™ TSSP

The Voice Direct™ TSSP provides cost-sensitive products with speaker-dependent speech recognition, speech synthesis and DTMF tone generation. This easy-to-use, pin-configurable chip requires no custom programming and can recognize up to 60 trained words. The Voice Direct™ TSSP is most ideal for consumer telephony products which feature voice dialing.

Voice Dialer™ ASSP

The Voice Dialer™ ASSP delivers speech recognition technology that allows users to dial phone numbers by saying the name of the person they wish to call. Voice dialing and phone directory management through speech recognition can be easily integrated into existing products. This IC is designed for use as a slave chip controlled by an external host processor.

Voice Dialer Software

The Voice Dialer software provides advanced speech technology on a variety of microcontroller and DSP platforms. A complete speech API and flexible design allows manufacturers to easily integrate speech functionality into telephony products.

IMPORTANT NOTICES

Sensory reserves the right to make changes to or to discontinue any product or service identified in this publication at any time without notice in order to improve design and supply the best possible product. Sensory does not assume responsibility for use of any circuitry other than circuitry entirely embodied in a Sensory product. Information contained herein is provided gratuitously and without liability to any user. Reasonable efforts have been made to verify the accuracy of this information but no guarantee whatsoever is given as to the accuracy or as to its applicability to particular uses.

Applications described in this data sheet are for illustrative purposes only, and Sensory makes no warranties or representations that the RSC series of products will be suitable for such applications. In every instance, it must be the responsibility of the user to determine the suitability of the products for each application. Sensory products are not authorized for use as critical components in life support devices or systems.

Sensory conveys no license or title, either expressed or implied, under any patent, copyright, or mask work right to the RSC series of products, and Sensory makes balance between recognition and synthesis no warranties or representations that the RSC series of products are free from patent, copyright, or mask work right infringement, unless otherwise specified. Nothing contained herein shall be construed as a recommendation to use any product in violation of existing patents or other rights of third parties. The sale of any Sensory product is subject to all Sensory Terms and Conditions of Sales and Sales Policies.



S E N S O R Y

521 East Weddell Drive
Sunnyvale, CA 94089

TEL: (408) 744-9000
FAX: (408) 744-1299

© 1996 SENSORY, INC.
ALL RIGHTS RESERVED
P/N 80-0022-5 9/8/98

Sensory is registered by the U.S. Patent and Trademark Office.
All other trademarks or registered trademarks are the property
of their respective owners.

From the *Interactive Speech™* Line of Products