

# **VOICE DIRECT™ 364**

## **Quick Setup Guide**

### **CONTENTS OF THIS GUIDE:**

1. *Introduction*
2. *Features*
3. *Building a Basic Application*
4. *Module Pinout and Characteristics*
5. *Training VOICE DIRECT™ 364*
6. *Recognizing Words*
7. *Erasing Templates*
8. *Tips for Optimal Performance*
9. *About Speech Recognition*
10. *Module Schematic*
11. *Bill of Materials*
12. *How to learn more about VOICE DIRECT™ 364*
13. *Other Sensory Products*
14. *Disclaimers and Warranty*

### **1. Introduction**

This guide will walk you through the steps of bringing speech recognition to your product or application. First, we'll focus on getting you up and running quickly. Later on, we'll discuss tips, tricks, and other information about Sensory speech recognition technology. For those interested in detailed technical information about the VOICE DIRECT™ 364 module and Sensory's award winning speech recognition technology, please visit our web site at [\*\*www.VoiceActivation.com\*\*](http://www.VoiceActivation.com).

### **2. Features**

- Speaker-Dependent and Continuous Listening speech recognition technologies
- Minimal external components
- Recognizes up to 60 words or phrases in slave mode, or 15 in stand-alone mode (broken in to 1, 2 or 3 sets)
- Over 99% recognition accuracy with proper design
- Phrase recognition up to 2.5 seconds
- User-friendly speech prompting

### **3. Building a Basic Application**

There are a few items that should be considered before starting construction:

#### **CONSTRUCTION**

Consider how you are going to fixture the module. Whether or not you mount the module in a box, for example, depends on how much additional circuitry you'd like to add to your application, and how transportable it needs to be.

- ◆ One possible method of construction is to use a small project box, mounting the speaker, microphone, switches, and module (on standoffs) in the box. For connections, 30 AWG wire-wrap wire can be soldered directly from the top (non-connector side) of the module to the other components.
- ◆ An alternate method of construction is to use a piece of perf or proto board (approx. 2"x5"), mounting male header strips (0.1" centers) to the perf board to accommodate the module. The perf board can also be used to mount the switches, as well as any other components required by your application. 1/8" mono phone jacks and plugs are useful for quickly connecting and disconnecting the speaker, microphone and power from the rest of the circuitry.

#### **POWER**

Three 1.5V batteries in a plastic holder, or a lab supply, can be used as a power source. A power switch in series with one of the power leads will maximize battery life by allowing you to power down the module when not in use. Incorporating a protection diode in series with the power supply will avoid damage to the circuit if batteries are inserted with the wrong polarity.

#### **OUTPUTS**

- ◆ Consider what types of loads you are going to be driving with the outputs. Most types of loads (LEDs, motors, relays, etc.) will require some type of output driver circuitry. A discussion of these circuits is beyond the scope of this guide, but a great place to start looking is in the databooks, application notes, and on the web pages of major semiconductor manufacturers.
- ◆ Keep in mind that the outputs only toggle high for 1 second. If you want them to stay high, you can add a latching output driver. The outputs also toggle as part of the power up initialization sequence, so bear in mind how that may affect your application.
- ◆ Because VOICE DIRECT™ 364 can recognize 15 words in stand-alone mode, but has only 8 outputs, some decoding circuitry is required if the application requires more than 8 words to be recognized. The format of the outputted words is discussed under "Recognizing Words".

## SYSTEM DESIGN

- ◆ If the module is used in a system with other digital clocks (switching power supplies, LCD driver, etc.) take special care to prevent these signals from being coupled into the audio circuitry of VOICE DIRECT™ 364.
- ◆ With proper product construction, Voice Direct™ 364 meets the CE requirements for electromagnetic radiation and immunity. To minimize radiated emissions, speaker wires should be less than 3" long. In addition, the speaker cable and power cable should be oriented on opposite sides of the module.

## MICROPHONE CONSIDERATIONS

For most applications, an inexpensive omni-directional electret capacitor microphone with a minimum sensitivity of -60 dB is adequate. In some applications, a directional microphone might be more suitable if the signal comes from a different direction than the audio noise. Since directional microphones have a frequency response that depends on their distance from the sound source, such microphones should be used with caution. For best performance, speech recognition products should be used in a quiet environment with the speaker's mouth in close proximity to the microphone.

Important mechanical issues pertaining to microphone assembly:

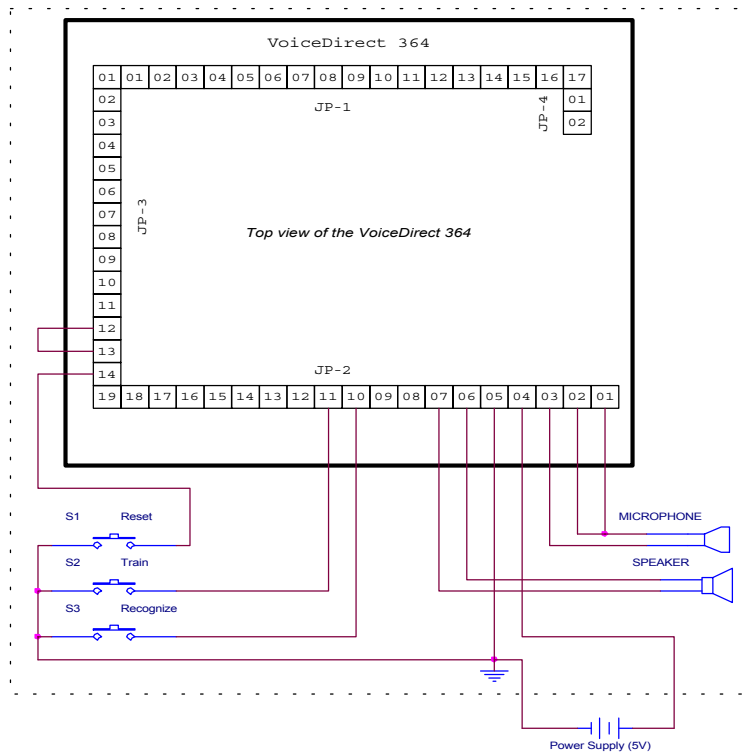
- ◆ FIRST: In the product, the microphone element should be positioned as close to the mounting surface as possible and should be fully seated in any housing. There must be NO airspace between the microphone element and the housing. Having such an airspace can lead to acoustic resonance, which can reduce recognition accuracy.
- ◆ SECOND: The area in front of the microphone element must be kept clear of obstructions to avoid interference with recognition. In general, the diameter of the hole in the housing in front of the microphone should be at least 5 mm. Any necessary plastic surface in front of the microphone should be as thin as possible, being no more than 0.7 mm if possible.
- ◆ THIRD: The microphone should be acoustically isolated from the housing if possible. This can be accomplished by surrounding the microphone element with a spongy material such as rubber or foam. Mounting with a pliable, non-hardening adhesive is another possibility. The purpose is to prevent auditory noises produced by handling or jarring the product from being "picked up" by the microphone. Such extraneous noises can reduce recognition accuracy.

## MODE SELECTION

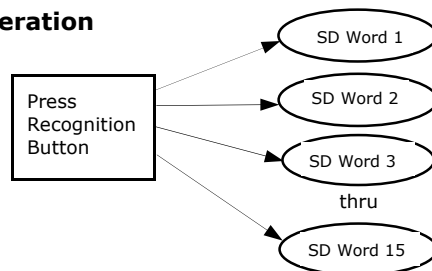
There are 2 main configuration modes for the module that can be selected depending on how voice recognition will interact with your application. These are Speaker-Dependent (SD) mode and Continuous Listening (CL) Mode. There are 2 types of CL modes: single word and multi-word. Following are schematics depicting how to configure the module for each mode. Operating instructions can be found in sections 4, 5 and 6 of this manual.

### 1. SD Configuration

**Figure 1**



### SD Operation

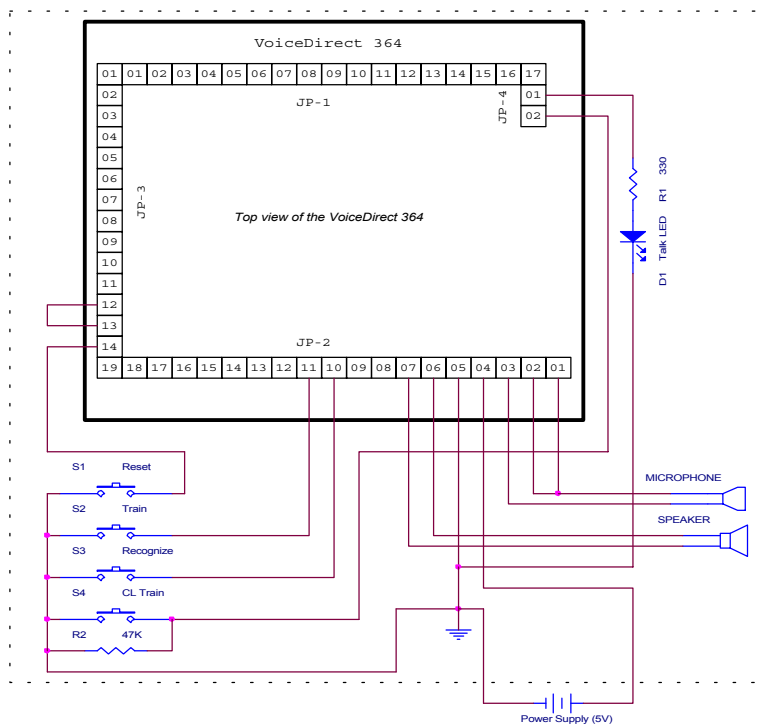


## 2. Continuous Listening (CL) Mode Configuration

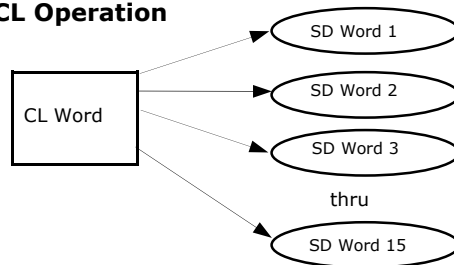
There are 2 separate configurations of CL mode:

### Single Word CL Configuration

Figure 2

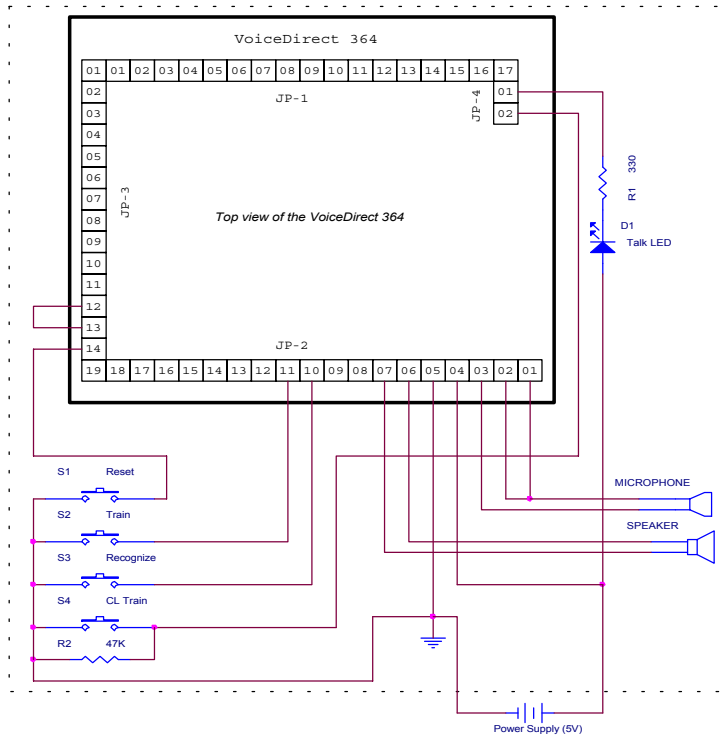


### Single Word CL Operation

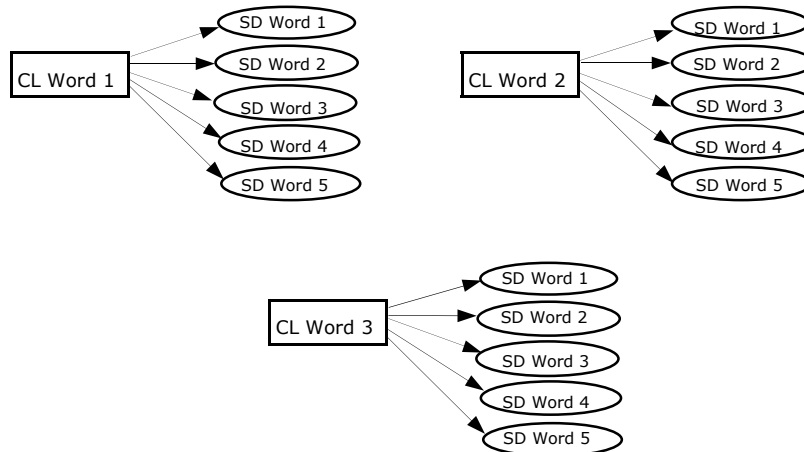


## Multi-word CL Configuration

**Figure 3**



## Multi-word CL Operation



## 4. Module Pinout and Characteristics

Pin #	Name	Description	I/O
<b>JP1:</b>			
01 ~ 17	-	Not Used – Leave open.	I
<b>JP2:</b>			
01	MIC IN	Microphone Input Connection	I
02	MIC BIAS	Microphone Bias (Electret microphone)	O
03	MIC RETURN	Analog Ground. For noise reasons, analog and digital grounds should connect together only at the Voice Direct 364	-
04	+5V	5 Volt (+) power supply connection	I
05	GND	Ground connection	I
06	SPEAKER (+)	Speaker connection (positive)	O
07	SPEAKER (-)	Speaker connection (internally connected to ground)	O
08	DACOUT	Analog output (unbuffered)	O
09	-MHS (P0.0)	Stand Alone Mode: Not Used Slave Mode: Serial Interface Master Handshake (MHS)	- I
10	-RECOG (P0.1)	Stand Alone Mode: Recognition Sensitivity (see below) Slave Mode: Not Used	I -
11	-TRAIN (P0.2)	Stand Alone Mode: Training Sensitivity (see below) Slave Mode: Not Used	I -
12	OUT1 (P1.0)	Stand Alone Mode: Output 1 or 9 (Active high) Slave Mode: Talk	O O
13	OUT2 (P1.1)	Stand Alone Mode: Output 2 or 10 (Active high) Slave Mode: Configurable I/O #0	O I/O
14	OUT3 (P1.2)	Stand Alone Mode: Output 3 or 11 (Active high) Slave Mode: Configurable I/O #1	O I/O
15	OUT4 (P1.3)	Stand Alone Mode: Output 4 or 12 (Active high) Slave Mode: Configurable I/O #2	O I/O
16	OUT5 (P1.4)	Stand Alone Mode: Output 5 or 13 (Active high) Slave Mode: Configurable I/O #3	O I/O
17	OUT6 (P1.5)	Stand Alone Mode: Output 6 or 14 (Active high) Slave Mode: Configurable I/O #4	O I/O
18	OUT7 / -SHS (P1.6)	Stand Alone Mode: Output 7 or 15 (Active high) Slave Mode: Serial Interface Slave Handshake (SHS)	O I/O
19	OUT8 / -DATA (P1.7)	Stand Alone Mode: Output 8 (Active high) Slave Mode: Serial Interface Data (DATA)	O I/O
<b>JP3:</b>			
01 ~ 09	-	Not Used – Leave open.	-
10	ERROR / -PROCESS (P0.3)	Stand Alone Mode: Error LED Slave Mode: Low when processing a command	O O
11	GND	Ground	I
12	GND	Ground	I
13	MODE	Slave/Stand-alone mode selection	I
14	-RESET	Reset (Active low)	I
<b>JP4:</b>			
01	-TALK / CL CFG (P0.6)	Stand Alone Mode: TALK LED, Single-/Multi-CL config. port Slave Mode: Not Used	I/O -
02	-CL TRAIN (P0.7)	Stand Alone Mode: CL Training, CL performance config. port Slave Mode: Not Used	I -

### Training and Recognition Selectivity

The -TRAIN, -CL TRAIN and -RECOG pins control the selectivity and activation of the VOICE DIRECT™ 364. The training and recognition selectivity levels are set when the VOICE DIRECT™ 364 is first powered up (or reset) according to the following table:

Pin Configuration	Selectivity
-TRAIN open circuit	Relaxed SD Training – Easier to train, accepts more similar sounding words (fewer rejections).
-TRAIN pulled to GND with a 100K resistor	Strict SD Training – Harder to train, rejects more similar sounding words – may result in better recognition accuracy (fewer substitutions).
-CL TRAIN pulled to GND with a 47K resistor	Relaxed CL Training – Easier to train, accepts more similar sounding words (fewer rejections).
-CL TRAIN pulled to GND with a 680K Resistor	Strict CL Training – Harder to train, rejects more similar sounding words – may result in better recognition accuracy (fewer substitutions).
-RECOG open circuit	Relaxed SD and CL Recognition – Recognizes more words, may substitute the wrong words (fewer rejections).
-RECOG pulled to GND with a 100K resistor	Strict SD and CL Recognition – Recognizes fewer words, may reject trained words (fewer substitutions).

Some applications of Continuous Listening require maximum accuracy with the fewest false triggers. This configuration is appropriate for an application such as a light switch, where a false trigger may be more than a minor inconvenience. Further, the CL recognizer in such an application hears a great many phrases that are *not* trigger phrases, so there are many opportunities for false triggers and a high level of accuracy is desired. A consequence of this high accuracy is relatively slow responsiveness.

Other applications – games, for example – can accept occasional false triggers, but need fast response. In these applications the relative frequency of trigger phrases may be quite high, so false triggering accuracy is less of a concern.

The tradeoff between CL accuracy and responsiveness is controlled by selection of R2. For slow CL performance, select a 680 K-Ohm resistor for R2. For fast CL performance, select a 47 K-Ohm resistor.



## Module Pin Characteristics

### Absolute Maximum Ratings

Any pin to GND	-0.1V to +6.5V	<b>WARNING:</b> <i>Stressing the Voice Direct™ 364 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</i>
Operating temperature(T <sub>0</sub> )	-20°C to +70°C	
Power dissipation	1 W	
Operating Conditions	-20°C to +70°C; V <sub>DD</sub> =4.5 - 5.25V V <sub>SS</sub> =0V	

### DC Characteristics

(T<sub>0</sub> = -20°C to +70°C, V<sub>dd</sub> = 5V )

SYMBOL	PARAMETER	MIN	TYP	MAX	UNITS	TEST CONDITIONS
V <sub>IL</sub>	Input Low Voltage ♦ MODE, I/O pins,	-0.1		0.75	V	
V <sub>IH</sub> (V <sub>dd</sub> >4.5)	Input High Voltage ♦ MODE, I/O pins,	3.0		V <sub>dd</sub> +0.3	V	
V <sub>OL</sub>	Output Low Voltage		0.3	0.1*V <sub>dd</sub>	V	I <sub>OL</sub> = 2 mA
V <sub>OH</sub>	Output High Voltage	0.8*V <sub>dd</sub>	0.9*V <sub>dd</sub>		V	I <sub>OL</sub> = -2 mA
I <sub>IL</sub>	Logical0 Input Current		<1	10	uA	V <sub>ss</sub> <V <sub>pin</sub> <V <sub>dd</sub>
I <sub>DD1</sub>	Supply Current, Operating		37	100	mA	Hi-Z Outputs
I <sub>DD2</sub>	Supply Current, Quiescent		4	6	mA	Hi-Z Outputs
R <sub>pu</sub>	Pull-up resistance, P0.0-P1.7	5, 80, Hi-Z	6.5, 200, Hi-Z		K-ohm	Selected with software

## **5. Training VoiceDirect™ 364**

### **SD Mode**

The module should be configured for SD Mode (Figure 1). When the –TRAIN pin is pulled to GND for at least 100mS (such as by pushing a momentary switch, labeled in the schematics as “TRAIN”), training will begin. VOICE DIRECT™ 364 will prompt “Say word x” (where x corresponds to the word to be trained).

A trained word or phrase must be shorter than 2.5 seconds and may not contain silences longer than 0.5 seconds. For example, the name “John Smith” would be an acceptable phrase as long as the two words are not separated by a large pause. Training terminates when no word is spoken in response to a prompt, when either switch is pressed a second time during training, when three errors have occurred during training, or after all words have been trained.

Pressing the TRAIN switch at a later time resumes training. New words are added to the end of the set already recorded. New words may be added to the set at any time, up to a maximum of 15 words. Individual words from the set may not be deleted or overwritten, but the entire set can be erased (see “Erasing Templates”, section 7). The user says the first word to be trained, then VOICE DIRECT™ 364 prompts again with “Repeat”. The user repeats the word, and VOICE DIRECT™ 364 will return “Accepted” if the word has been successfully trained, otherwise it will say the cause of the training error. If an error occurs during training, then the error will be spoken “Spoke too soon”, “Please talk louder”, etc., and the ERROR line will pulse high for 1 second. The user will get three attempts to train each word before VOICE DIRECT™ 364 exits training mode, says “Training complete”, and pulses the ERROR line high for 1S.

The user can exit training at any time by pressing the –TRAIN or the –RECOG button, by not responding to a “Say word x” or “Repeat” prompt, or when all 15 words have been trained.

**Example (SD):**

<b>Action:</b>	<b>Press TRAIN switch</b>	<b>Start training</b>
Voice Direct 364:	"Say word one"	<i>Training first SD word</i>
User:	"Start"	
Voice Direct 364:	"Repeat"	
User:	"Start"	
Voice Direct 364:	"Say word two"	<i>Training second SD word</i>
User:	"Shutdown"	
Voice Direct 364:	"Repeat"	
User:	"Shutdown"	
Voice Direct 364:	"Say word three"	<i>Training third SD word</i>
User:	"Lights"	
Voice Direct 364:	"Repeat"	
User:	"Lights"	
<b>Action:</b>	<b>Press TRAIN switch</b>	<b>Stop training</b>

**Note:** Voice Direct 364 will continue to prompt for new words to train until all 15 memory locations have been filled. To stop training, press the TRAIN switch again.

### Single Word CL Mode

The module should be configured for Single Word CL Mode (Figure 2). In this mode, separate buttons are used to train SD words (TRAIN switch) and the CL word (CL TRAIN switch). After training the CL word, up to 15 SD commands can be trained. Another difference from SD mode is that while training SD words, the TRAIN switch must be pressed to train each word.

#### **Example (Single Word CL):**

<b>Action:</b>	<b>Press CL TRAIN switch</b>	<b>Start CL training</b>
Voice Direct 364:	"Say word one"	Training CL word
User:	"Sound System"	
Voice Direct 364:	"Repeat"	
User:	"Sound System"	
<b>Action:</b>	<b>Press TRAIN switch</b>	<b>Start SD training</b>
Voice Direct 364:	"Say word one"	Training first SD word in the set
User:	"Start"	
Voice Direct 364:	"Repeat"	
User:	"Start"	
<b>Action:</b>	<b>Press TRAIN switch</b>	<b>Continue SD training</b>
Voice Direct 364:	"Say word two"	Training second SD word in the set
User:	"Shutdown"	
Voice Direct 364:	"Repeat"	
User:	"Shutdown"	
<b>Action:</b>	<b>Press TRAIN switch</b>	<b>Continue SD training</b>
Voice Direct 364:	"Say word three"	Training third SD word in the set
User:	"Lights"	
Voice Direct 364:	"Repeat"	

## Multi-Word CL Mode

The module should be configured for Multi-Word CL Mode (Figure 3). The training sequence is very similar to Single Word CL mode, except that 3 separate CL words can be trained. After training each CL command, up to 5 SD words can be trained. ***It is important to note that each SD set must be trained immediately following the corresponding CL word. Once the next CL word is trained, there is no way to go back and re-train an SD set without erasing the entire memory.*** Trying to train 2 similar CL words will result in a "too similar" error. During SD training, each SD word is checked for similarity to other words in its set, but not against those in other sets.

Multi-Word CL mode can be useful for a single user product using multiple gateway commands, or for multi-user applications where each user would train his or her own CL and SD words.

### **Example (Multi-Word CL):**

<b>Action:</b>	<b>Press CL TRAIN switch</b>	<b>Start CL training for User 1</b>
Voice Direct 364:	"Say word one"	<i>Training first CL word</i>
User:	"Sound System"	
Voice Direct 364:	"Repeat"	
User:	"Sound System"	
<b>Action:</b>	<b>Press TRAIN switch</b>	<b>Start SD training for User 1</b>
Voice Direct 364:	"Say word one-one"	<i>Training first SD word in first set</i>
User:	"Start"	
Voice Direct 364:	"Repeat"	
User:	"Start"	
<b>Action:</b>	<b>Press TRAIN switch</b>	<b>Continue SD training for User 1</b>
Voice Direct 364:	"Say word one-two"	<i>Training second SD word in first set</i>
User:	"Shutdown"	
Voice Direct 364:	"Repeat"	
User:	"Shutdown"	
<b>Action:</b>	<b>Press CL TRAIN switch</b>	<b>Start CL training for User 2</b>
Voice Direct 364:	"Say word one"	<i>Training second CL word</i>
User:	"Sound System"	
Voice Direct 364:	"Repeat"	
User:	"Sound System"	

### Example (Multi-Word CL) continued:

<b>Action:</b>	<b>Press TRAIN switch</b>	<b><i>Start SD training for User 2</i></b>
Voice Direct 364:	"Say word two-one"	<i>Training first SD word in second set</i>
User:	"Start"	
Voice Direct 364:	"Repeat"	
User:	"Start"	
<b>Action:</b>	<b>Press TRAIN switch</b>	<b><i>Continue SD training for User 2</i></b>
Voice Direct 364:	"Say word two-two"	<i>Training second SD word in second set</i>
User:	"Shutdown"	
Voice Direct 364:	"Repeat"	
User:	"Shutdown"	

## 6. Recognizing Words

### SD Mode

When the -RECOG pin is pulled to GND for at least 100mS, recognition will begin. VOICE DIRECT™ 364 will prompt "Say a word". If the response is not recognized, VOICE DIRECT™ 364 will say "Word not recognized" and exit recognition mode. If the spoken word matches a stored template, one (or two) of 8 output pins is activated (pulled high for approximately one second) and a voice message indicates the matching response. If the set contains 8 or fewer elements, these pins may be used to control actions directly. If the set contains more than 8 elements, decoding is necessary. The logical format of the outputs is shown in the table below.

Recognition Word	OUT 1	OUT 2	OUT 3	OUT 4	OUT 5	OUT 6	OUT 7	OUT 8
SD Word 01	A							
SD Word 02		A						
SD Word 03			A					
SD Word 04				A				
SD Word 05					A			
SD Word 06						A		
SD Word 07							A	
SD Word 08								A
SD Word 09	A							A
SD Word 10		A						A
SD Word 11			A					A
SD Word 12				A				A
SD Word 13					A			A
SD Word 14						A		A
SD Word 15							A	A

**Note: "A" indicates that the outputs are "Active-high".**

If an error occurs during recognition (except for "Word not recognized"), then the error will be spoken "Spoke too soon", "Please talk louder", etc., and the ERROR line will pulse high for 1 second. If the spoken word is not recognized for any reason, none of the pins is activated and an appropriate voice message is synthesized. Since this can occur if VOICE DIRECT™ 364 has not been trained on the word, the "word not recognized" output is not considered an error condition, and the ERROR bit remains inactive.

For each SD word, VOICE DIRECT™ 364 monitors the background noise level and gives a voice warning if the noise is too high for recognition. VOICE DIRECT™ 364 works well with high (above 80 dB) levels of *steady* background noise (such as a fan), but it may make errors at lower levels if the background noise is not steady (for example, a TV set). Best results will be obtained in a relatively quiet location.

Warnings may also appear if the word is spoken too softly, too loudly, or too quickly after the prompt. The ERROR signal goes high when these conditions occur.

### Single Word CL Mode

Upon power-up or reset, VOICE DIRECT™ 364 checks to see if it has been configured in a Continuous Listening mode. If any words have been trained, VOICE DIRECT™ 364 will start listening for the CL word. After training, the RECOGNITION button can be pressed to initiate continuous listening operation.

It is important to note that the module is only listening when the TALK LED is lit. Use the LED as an indicator of when you can speak commands, especially when first using the module. After successful recognition of the CL word, there is a 3 second window during which VOICE DIRECT™ 364 is listening for an SD word (if any are trained). If no SD words are trained, then a successful recognition of the CL word will cause the outputs to behave as if the first SD word was recognized (OUT1 would toggle high). This feature is useful if only a single command is required, such as a light switch. Note that VOICE DIRECT™ 364 does not need to wait the full 3 seconds if it hears your command once it detects the silence at the end of your utterance, recognition processing will begin immediately.

When a trained sequence of words (CL word + SD word) is recognized, the appropriate Output Pin(s) will pulse high for 1 second:

Recognition Word	OUT 1	OUT 2	OUT 3	OUT 4	OUT 5	OUT 6	OUT 7	OUT 8
CL + SD Word 01	A							
CL + SD Word 02		A						
CL + SD Word 03			A					
CL + SD Word 04				A				
CL + SD Word 05					A			
CL + SD Word 06						A		
CL + SD Word 07							A	
CL + SD Word 08								A
CL + SD Word 09	A							A
CL + SD Word 10		A						A
CL + SD Word 11			A					A
CL + SD Word 12				A				A
CL + SD Word 13					A			A
CL + SD Word 14						A		A
CL + SD Word 15							A	A

**Note: "A" indicates that the outputs are "Active-high".**



### Multi-Word CL Mode

Multi-Word CL mode works in a similar fashion to Single Word mode, except that the module is continuously monitoring for up to 3 words at the same time.

The word (SD) following the triggering word (CL) will be matched only against the templates in the corresponding set. If no SD words are trained, then a successful recognition of one of the CL words will cause the outputs to behave as if the first SD word in that set was recognized. For example, if no SD words were trained for Set B, then OUT2 and OUT4 would toggle high. This feature is useful in applications requiring up to 3 commands, or where there are multiple users.

CL Word	SD Word	OUT 1	OUT 2	OUT 3	OUT 4	OUT 5	OUT 6	OUT 7	OUT 8
CL Word A	SD Word 01	A			A				
CL Word A	SD Word 02	A				A			
CL Word A	SD Word 03	A					A		
CL Word A	SD Word 04	A						A	
CL Word A	SD Word 05	A							A
CL Word B	SD Word 01		A		A				
CL Word B	SD Word 02		A			A			
CL Word B	SD Word 03		A				A		
CL Word B	SD Word 04		A					A	
CL Word B	SD Word 05		A						A
CL Word C	SD Word 01			A	A				
CL Word C	SD Word 02			A		A			
CL Word C	SD Word 03			A			A		
CL Word C	SD Word 04			A				A	
CL Word C	SD Word 05			A					A

**Note:** "A" indicates that the outputs are "Active-high".

## **7. Erasing the Templates**

When both the –TRAIN and –RECOG pin are pulled to GND for at least 100mS, VOICE DIRECT™ 364 will erase all trained templates. VOICE DIRECT™ will say “Memory erased”, pulse the ERROR line high for 1S, and then exit the erase mode. Individual words or sets cannot be erased separately.

## **8. Tips for Optimal Performance**

Successful recognition begins with the careful selection of words for each recognition set. Several factors contribute to selecting an optimal recognition set. Problematic recognition sets can often be corrected by replacing one or more words with a synonym, or approximate synonym (see examples, below), without requiring any other changes. The smaller the set, the higher the recognition rate.

<b>The optimal set consists of:</b>	<b>Avoid sets like:</b>	<b>Aim for sets like:</b>
Dissimilar sounding words	hat/cat/rat home phone/office phone	hat/kitten/mouse home/office
Varying numbers of syllables	orange/apple/cherry	orange/watermelon/ grape

### ***Key considerations for successful voice recording***

- The equipment used to train the voice recordings should match the equipment used during recognition. Differences in microphone, microphone housing, etc. will adversely affect recognition.
- The conditions and environment in which the voice recordings are made should reflect the conditions and environment in which the end product will be used.
  - a) **Distance.** The distance of the microphone from the speaker's mouth must be the same during recording and during end-product use. For example, a doll is typically held within arm's length, and so the voice recording microphone should be held accordingly.
  - b) **Natural Voices.** Subjects should speak in their normal voice and should be discouraged from sounding different by imitating a foreign accent or using any unnatural intonation. They should be prompted by means of some non-verbal source (pictures or flashcards, for example), so as not to unconsciously mimic the voice of the person doing the prompting.
  - c) **Physical States.** Physical states should be considered. For example, in collecting voice recordings for an exercise machine, it is strongly recommended to record people who are out of breath.
  - d) **Emotional States.** Emotional states should be considered. Will the end users be relatively quiet and calm (say, for an office product) or loud and excited (say, for a playground toy)?
  - e) **Environment / Background Noise.** Environmental noise must be considered. Voice recordings should ideally be made in an environment similar to the one in which the end product will be used. For a product intended for use in cars, it is advisable to record in a running, moving car. Nevertheless, the recordings must be made in a reasonably quiet environment. The speech signal must be prominent relative to background noise and there should not be any abrupt, loud noises. Voice recordings should not be made in a soundproof room. These rooms lend an unnatural background silence to the recordings, which does not reflect the real-world environment in which the end product will be used.

## 9. About Speech Recognition

VOICE DIRECT™ 364 performs speaker-dependent discrete word recognition by comparing a pattern generated in real time with previously trained word templates. The pattern generated by VOICE DIRECT™ 364 is based on a digital reconstruction of the voice command. Each word to be recognized must first be *trained*. During training, VOICE DIRECT™ 364 constructs a *template* representing the individual speaker's unique sound pattern for each specific word or phrase to be recognized. Templates are stored in serial EEPROM memory. During recognition a new pattern is produced and compared to the stored templates to determine which word was spoken. VOICE DIRECT™ 364 features integrated speech prompting for both training and recognition operations, allowing the development of sophisticated interactive products with minimal programming. VOICE DIRECT™ 364 performs the following operations when recognizing a word:

1. The audio signal (spoken word) is externally amplified and filtered and then supplied to the analog inputs of the VOICE DIRECT™ 364, which converts the analog waveforms to digital samples.
2. VOICE DIRECT™ 364 analyzes the speech signal samples and generates a pattern of information representing significant speech elements.
3. VOICE DIRECT™ 364 increases or decreases the gain of the external amplifier as needed to maintain signal quality.
4. Using a neural network, the pattern is compared with previously stored template patterns; a small number of candidate templates is selected.
5. The candidate templates are further processed to determine the one template that provides the best match to the unknown pattern.
6. If the best match template gives a score above a pre-defined threshold, VOICE DIRECT™ 364 chooses the word associated with that template. If no template provides a match above threshold, a special "no match" value is chosen.

Steps 1-3 above are repeated for each word during training. VOICE DIRECT™ 364 stores the average of two training patterns for each word to improve accuracy. Before storing a new template it is compared with the existing templates in the set. The new candidate word will not be accepted if it is too similar to an existing word (for example, "Bill Smith" and "Jill Smith").

VOICE DIRECT™ 364, like other speech recognition systems, is necessarily subject to two types of errors: *rejects* (failure to recognize a word in the vocabulary) and *substitutions* (confusion of two vocabulary words, or recognition of a non-vocabulary word). The relative importance of each type of error may depend on the application. VOICE DIRECT™ 364 provides selectivity levels that allow the user to optimize the tradeoff between these two types of recognition errors. When the recognition selectivity level is set to its highest value, VOICE DIRECT™ 364 minimizes substitution errors but may produce more rejection errors. When the training selectivity level is set to its highest value, VOICE DIRECT™ 364 minimizes both substitution and rejection errors by rejecting words that are too similar sounding, potentially increasing recognition accuracy. These settings are best established by experimenting with your own application and environment.



*This page intentionally left blank.*

## 11. Bill of Materials for VOICE DIRECT™ 364

Item	Qty	Part Title	Reference Designator
1	1	PCB, Voice Direct 364 PCB	
2	1	EPROM, Voice Direct 364, English	U2*
3	1	TQFP 10x10x1.4, RSC-300	U1
4	2	Res, 100K	R6,R7*,R12
5	1	Res, 10	R19
6	2	Res, 1K	R4,R16
7	2	Res, 100	R8,R9
8	1	Res, 2.7K	R1
9	1	Res, 4.7K	R11
10	2	Res, 47K	R15,R18
11	1	Res, 220K	R2
12	2	Res, 300K	R14,R17
13	1	Res, 0	R5,R10*
14	1	Res, 47	R3
15	1	Res, 1.2M	R13
16	1	Cap, 27pF / 50V	C11
17	1	Cap, 100pf / 50V	C20
18	1	Cap, 1uF / 16V	C16
19	5	Cap, 100uF / 6.3V	C5,C8,C15,C17,C19
20	9	Cap, 0.1uF / 50V	C1,C3,C9,C10,C12,C13,C14,C21,C22
21	1	Cap, 0.01uF / 50V	C6
22	1	Cap, 220pF / 50V	C4
23	1	Cap, 0.022uF / 50V	C7
24	1	Cap, 4700pF / 100V	C2
25	1	Cap, 1000 pF / 50V	C18
26	1	24LC65T/SM	U3
27	1	LM386M-1	U4
28	1	Header, 17 Pin Female	JP1
29	1	Header, 19 Pin Female	JP2
30	1	Header, 14 Pin Female	JP3
31	1	Header, 2 Pin Female	JP4
32	1	Crystal, 14.3 MHz	Y1

\* **Note:** Some Voice Direct 364 modules contain firmware built into U1. For these modules, U2 is not included, R6 is placed in position R7, and R5 is placed in position R10.



## 12. How to learn more about VOICE DIRECT™ 364

Voice Direct™ 364 has more advanced capabilities designed for embedded systems developers. Please note that user-defined synthesis is not currently supported for Voice Direct™ 364—contact Sensory for custom development services. Available as a single chip in a 64 lead 10x10mm TQFP package or in die form, Voice Direct™ 364 is priced competitively for low to high-volume applications. For a complete databook that includes detailed technical specifications, including programming information on using Voice Direct™ 364 as a slave-mode voice recognition processor, visit our website at **www.VoiceActivation.com**.

## 13. Other Sensory Products

### 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 **VoiceActivation™ Software** technologies run on a variety of microcontrollers and DSPs.

### RSC Microcontrollers

The RSC family of microcontrollers (RSC-164, RSC-200/264T, and RSC-300/364) are low-cost 8-bit microcontrollers designed for use in consumer electronics. All members of the RSC family are fully integrated and include a speech processor, A/D, D/A, ROM (except RSC-200/300), and RAM circuitry on chip. The RSC-200/264T and RSC-300/364 also include on chip pre-amplification. The RSC family of microcontrollers can perform a full range of speech/audio functions including speech recognition, speaker verification, speech and music synthesis, and voice record/playback.



**Voice Dialer™ 364 ASSP**

The Voice Dialer™ 364 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. The Voice Dialer™ 364 product line includes IC's and a low cost development kit.

**Voice Activation™ Software**

Sensory's Voice Activation™ 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.

## ***14. Disclaimers and Warranty***

**WARNING**

This kit is intended for use by consumers experienced with building electronic kits. As with any electronic kit, caution should be exercised during assembly, and all connections should be double-checked that they are clean, safe and properly soldered before applying any power source.

**Important Disclaimer**

To the fullest extent permitted by acceptable law, Sensory, Inc. expressly disclaims the implied warranty of fitness for a particular purpose. Customer should understand that Sensory does not make any representation that products purchased will suit customer's particular purpose. Customer must rely on customer's own skill or judgement in selecting suitable products for customer. To the extent any implied warranty is applicable, such warranty shall be limited to 90 days from the date of purchase.

## **Voice Direct™ 364 Speech Recognition Kit Limited Warranty**

The Voice Direct 364 Speech Recognition Kit is warranted against defects and workmanship for a period of 90 days from the date of product purchase. Sensory, Inc. will, at its option, either repair or replace a product that proves to be defective either upon receipt or through normal usage. If a Sensory Speech Recognition Kit product has been obsoleted or is no longer in production and deemed non-repairable, Sensory will, at its option, provide an equivalent product or system for a nominal fee.

Sensory, Inc. warrants this Speech Recognition Kit product, when properly installed and used, will execute its programmed instructions. However, Sensory, Inc. does not warrant that the operation of the Product, its firmware and software will be uninterrupted or totally error free. The Product must be returned to Sensory, Inc. for warranty service within the warranty period to the following address: Sensory, Inc., 521 E. Weddell Drive, Sunnyvale, CA 94089-2164. The Buyer will pay all shipping and other charges or assessments for the return of the Product to Sensory, Inc.

### **Limitation of Warranty**

The foregoing warranty shall not apply to defects resulting from maintenance performed by anyone other than Sensory, Inc., modifications made by Buyer or any third party, Buyer supplied software or interfacing, misuse, abuse, accident, mishandling, operation outside the environmental specifications for the Product, or improper setup or maintenance.

### **Limitation of Liability**

Sensory's liability shall be limited to the repair or replacement of defective products in accordance with the Voice Direct Speech Recognition Kit Limited Warranty. Sensory shall not be liable for any incidental, special or consequential damages for breach of any warranty, expressed or implied, directly or indirectly arising out of Sensory's sale of merchandise, including any failure to deliver any merchandise, or arising out of customer's installation or use, whether proper or improper, of the product, separately or in combination with other equipment, or from any other cause. Products sold by Sensory are not authorized for use as critical components in life support devices or systems.

### **Exclusive Remedies**

The remedies provided herein are Sensory's sole liability and Buyer's sole and exclusive remedies for breach of warranty. Sensory shall not be liable for any special, incidental, consequential, direct or indirect damages, whether based on contract, tort, or any legal theory. The foregoing warranty is in lieu of any and all other warranties, whether express, implied, or statutory, including but not limited to warranties of merchantability and suitability for a particular purpose.

### **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.

**Document 80-0180-3**