Example (SCL):

Action:	**Bring –TRIG TRAIN pin low**	***Start trigger word training***
Voice Direct™ II:	"Say word one"	*Training trigger word*
User:	"Sound System"	
Voice Direct™ II:	"Repeat"	
User:	"Sound System"	
Action:	**Bring -TRAIN pin low**	***Start command words training***
Voice Direct™ II:	"Say word one"	*Training first command word*
User:	"Start"	
Voice Direct™ II:	"Repeat"	
User:	"Start"	
Voice Direct™ II:	"Say word two"	*Training second command word*
User:	"Shutdown"	
Voice Direct™ II:	"Repeat"	
User:	"Shutdown"	
Voice Direct™ II:	"Say word three"	*Training third command word*
User:	"Lights"	
Voice Direct™ II:	"Repeat"	
User:	"Lights"	
Action:	**Bring -TRAIN pin low**	***Stop training***

Note: Voice Direct™ II will continue to prompt for new words to train until all 15 memory locations have been filled. To stop training, bring the -TRAIN pin low again.

MULTI-TRIGGER CONTINUOUS LISTENING (MCL) MODE
The module should be configured for MCL Mode. The training sequence is very similar to SCL mode, except that 3 separate trigger words can be trained. After training each trigger, up to 5 command words can be trained which are associated with that trigger word. ***It is important to note that each command set must be trained immediately following the corresponding trigger word. Once the next trigger word is trained, there is no way to go back and re-train a command set without erasing the entire memory.*** Trying to train two similar trigger words will result in a "too similar" error. During the training of the command set, each command word is checked for similarity to other command words in its set, but not against those in other sets.

MCL 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 trigger and command words.

Example (MCL):

Action:	**Bring –TRIG TRAIN pin low**	***Start set 1 trigger word training***
Voice Direct™ II:	"Say word one"	*Training set 1 trigger word*
User:	"Sound System"	
Voice Direct™ II:	"Repeat"	
User:	"Sound System"	
Action:	**Bring -TRAIN pin low**	***Start set 1 command words training***
Voice Direct™ II:	"Say word one-one"	*Training set 1 first command word*
User:	"Start"	
Voice Direct™ II:	"Repeat"	
User:	"Start"	
Voice Direct™ II:	"Say word one-two"	*Training set 1 second command word*
User:	"Shutdown"	
Voice Direct™ II:	"Repeat"	
User:	"Shutdown"	
Action:	**Bring -TRAIN pin low**	***Stop training***
Action:	**Bring –TRIG TRAIN pin low**	***Start set 2 trigger word training***
Voice Direct™ II:	"Say word one"	*Training set 2 trigger word*

User:	"Sound System"	
Voice Direct™ II:	"Repeat"	
User:	"Sound System"	
Action:	**Bring -TRAIN pin low**	***Start set 2 command words training***
Voice Direct™ II:	"Say word two-one"	*Training set 2 first command word*
User:	"Start"	
Voice Direct™ II:	"Repeat"	
User:	"Start"	
Voice Direct™ II:	"Say word two-two"	*Training set 2 second command word*
User:	"Shutdown"	
Voice Direct™ II:	"Repeat"	
User:	"Shutdown"	
Action:	**Bring -TRAIN pin low**	***Stop training***

Note: Voice Direct™ II will continue to prompt for new words to train until all 5 memory locations for each trigger word have been filled. To stop training, bring the -TRAIN pin low again.

SINGLE-TRIGGER WORD-SPOTTING MODE (SWS) MODE
The module should be configured for Word-Spotting Mode. In this mode, separate pins are used to train the trigger word (-TRIG TRAIN pin) and the command words (-TRAIN pin). After training the trigger word, up to 15 command words can be trained. The main difference between the two continuous listening modes (SCL and MCL) and single-trigger word-spotting (SWS) is that a continuous listening trigger word must be spoken by itself, with at least .25s of silence both before and after the word, while word-spotting trigger words may be embedded in the middle of a sentence., i.e. no before and after silence is needed.

Example (Single-Trigger Word-Spotting):

Action:	**Bring –TRIG TRAIN pin low**	***Start trigger word training***
Voice Direct™ II:	"Say word one"	*Training trigger word*
User:	"Sound System"	
Voice Direct™ II:	"Repeat"	
User:	"Sound System"	
Action:	**Bring -TRAIN pin low**	***Start command words training***
Voice Direct™ II:	"Say word one"	*Training first command word*
User:	"Start"	
Voice Direct™ II:	"Repeat"	
User:	"Start"	
Voice Direct™ II:	"Say word two"	*Training second command word*
User:	"Shutdown"	
Voice Direct™ II:	"Repeat"	
User:	"Shutdown"	
Voice Direct™ II:	"Say word three"	*Training third command word*
User:	"Lights"	
Voice Direct™ II:	"Repeat"	
User:	"Lights"	
Action:	**Bring -TRAIN pin low**	***Stop training***

Note: Voice Direct™ II will continue to prompt for new words to train until all 15 command words have been trained. To stop training, bring the -TRAIN pin low again.

8. Recognizing Words

EDGE-TRIGGERED SINGLE-RECOGNITION (ESR) MODE
When the –RECOG pin is pulled to GND for at least 100mS, recognition will begin. Voice Direct™ II will prompt "Say a word". If the response is not recognized, Voice Direct™ II will say "Word not recognized" and exit recognition mode. If the spoken word matches a stored template, one or two of 8 output pins are 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	OUT1	OUT2	OUT3	OUT4	OUT5	OUT6	OUT7	OUT8
Command Word 01	A							
Command Word 02		A						
Command Word 03			A					
Command Word 04				A				
Command Word 05					A			
Command Word 06						A		
Command Word 07							A	
Command Word 08								A
Command Word 09	A							A
Command Word 10		A						A
Command Word 11			A					A
Command Word 12				A				A
Command Word 13					A			A
Command Word 14						A		A
Command 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. If the spoken word is not recognized for any reason, none of the pins are activated and an appropriate voice message, "word not recognized", is synthesized.
For each command word, Voice Direct™ II monitors the background noise level. Voice Direct™ II works well with high (about 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.

SINGLE-TRIGGER CONTINUOUS LISTENING (SCL) MODE
When configured for SCL mode, the Voice Direct™ II will, upon power-up or reset, check to see if any words have been trained. If so, then Voice Direct™ II will immediately start listening for the trigger word. If no words have been trained, then the Voice Direct™ II will wait for the user to train one or more words. After training, the -RECOG pin can be brought low 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 the trigger and command words, especially when first using the module. After successful recognition of the trigger word, there is a 3 second window during which Voice Direct™ II is listening for a command word (if any are trained). If no command words are trained, then a successful recognition of the trigger word will cause the outputs to behave as if the first command word had been recognized (OUT1 would toggle high). This feature is useful if only a single trigger word is required. Note that Voice Direct™ II does not need to wait the full 3 seconds if it hears your command word 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:

I_{IL}	Logical 0 Input Current		<1	10	uA	$V_{ss}<V_{pin}<V_{dd}$
I_{DD1}	Supply Current, Operating		37	100	mA	Hi-Z Outputs
I_{DD2}	Supply Current, Quiescent		4	6	mA	Hi-Z Outputs
Rpu	Pull-up resistance, P0.0-P1.7	5, 80, Hi-Z	6.5, 200, Hi-Z		K-ohm	Selected with software

7. Training Voice Direct™ II

EDGE-TRIGGERED SINGLE-RECOGNITION (ESR) MODE
The module should be configured for SD Mode. When the –TRAIN pin is pulled to GND for at least 100mS (such as by pushing a momentary pushbutton, labeled in figure 2 as "S1 (Train)"), training will begin. Voice Direct™ II will prompt "Say word *x*" (where *x* corresponds to the word to be trained).

A trained word or phrase must be no longer 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 any pushbutton is pressed a second time during training, when three errors have occurred during training, or after all 15 words have been trained.

Bringing the -TRAIN pin low 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™ II prompts again with "Repeat". The user repeats the word, and Voice Direct™ II will return "Accepted" if the word has been successfully trained. Otherwise, it will indicate 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. The user will get three attempts to train each word before Voice Direct™ II exits training mode, and says "Training complete".

The user can exit training at any time by bringing the –TRAIN or the –RECOG pin low, by not responding to a "Say word *x*" or "Repeat" prompt, or when all 15 words have been trained.

Example (SD):

Action:	**Bring -TRAIN pin low**	***Start training***
Voice Direct™ II:	"Say word one"	*Training first command word*
User:	"Start"	
Voice Direct™ II:	"Repeat"	
User:	"Start"	
Voice Direct™ II:	"Say word two"	*Training second command word*
User:	"Shutdown"	
Voice Direct™ II:	"Repeat"	
User:	"Shutdown"	
Voice Direct™ II:	"Say word three"	*Training third command word*
User:	"Lights"	
Voice Direct™ II:	"Repeat"	
User:	"Lights"	
Action:	**Bring -TRAIN pin low**	***Stop training***

Note: Voice Direct™ II will continue to prompt for new words to train until all 15 memory locations have been filled. To stop training, bring the -TRAIN pin low again.

SINGLE-TRIGGER CONTINUOUS LISTENING (SCL) MODE
The module should be configured for SCL Mode. In this mode, separate pins are used to train the trigger word (-TRIG TRAIN pin) and the command words (-TRAIN pin). After training the trigger word, up to 15 command words can be trained.

Table 4 – Voice Direct™ II Sensitivity Threshold Configuration Pin Configuration	Selectivity
-TRAIN open circuit	Relaxed Training/Recognition – Easier to train, accepts more similar sounding words. Recognizes more words, may sometimes substitute the wrong words (fewer rejections).
-TRAIN pulled to GND with a 10K resistor	Strict Training/Recognition – Harder to train, rejects more similar sounding words. Recognizes fewer words, may sometimes reject trained words (fewer substitutions).
-TRIG TRAIN open circuit	Faster SCL/MCL/SWS Trigger Word Performance – Recognizes triggers more quickly, may sometimes substitute the wrong words (fewer rejections).
-TRIG TRAIN pulled to GND with a 10K resistor	Slower SCL/MCL/SWS Trigger Word Recognition – Recognizes triggers more slowly, may sometimes reject trained words (fewer substitutions).
-RECOG open circuit	Microphone sensitivity configured for up to arm's length optimal distance
-RECOG pulled to GND with a 10K resistor	Microphone sensitivity configured for up to 6 feet optimal distance

Some applications of Continuous Listening or Word-Spotting 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 SCL, MCL or SWS 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.

It should be noted that if any of the three pushbuttons S1, S2 or S3 are held down during reset or power-on, the Voice Direct™ II will interpret this as if the resistors are in circuit and will configure itself to use the strict setting for that pin.

MODULE PIN CHARACTERISTICS

Absolute Maximum Ratings

WARNING: *Stressing the Voice Direct™ II 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.*

Any pin to GND:	0.1V to +6.5V
Operating temperature(T_O):	20°C to +70°C
Power dissipation:	1 W
Operating Conditions:	20°C to +70°C; V_{DD}=2.7 – 3.6V V_{SS}=0V

DC Characteristics

(T_O = -20°C to +70°C, Vdd = 3.3V)

SYMBOL	PARAMETER	MIN	TYP	MAX	UNITS	TEST CONDITIONS
V_{IL}	Input Low Voltage P0.0-P1.7, -RESET	-0.1		0.75	V	
V_{IH}(V_{dd}>4.5)	Input High Voltage P0.0-P1.7, -RESET	3.0		Vdd + 0.3	V	
V_{OL}	Output Low Voltage P0.0-P1.7		0.3	0.1*Vdd	V	I_{OL}= 2 mA
V_{OH}	Output High Voltage P0.0-P1.7	0.8* Vdd	0.9*Vdd		V	I_{OL}= -2 mA

Recognition Word	OUT1	OUT2	OUT3	OUT4	OUT5	OUT6	OUT7	OUT8
Trigger + Command Word 01 (or Trigger + No Command Words Trained)	A							
Trigger + Command Word 02		A						
Trigger + Command Word 03			A					
Trigger + Command Word 04				A				
Trigger + Command Word 05					A			
Trigger + Command Word 06						A		
Trigger + Command Word 07							A	
Trigger + Command Word 08								A
Trigger + Command Word 09	A							A
Trigger + Command Word 10		A						A
Trigger + Command Word 11			A					A
Trigger + Command Word 12				A				A
Trigger + Command Word 13					A			A
Trigger + Command Word 14						A		A
Trigger + Command Word 15							A	A

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

MULTI-TRIGGER CONTINUOUS LISTENING (MCL) MODE

MCL mode works in a similar fashion to SCL mode, except that the module is continuously monitoring for up to three trigger words at the same time.

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

CL Word	SD Word	OUT1	OUT2	OUT3	OUT4	OUT5	OUT6	OUT7	OUT8
Trigger Word 1	Command Word 1	A			A				
Trigger Word 1	Command Word 2	A				A			
Trigger Word 1	Command Word 3	A					A		
Trigger Word 1	Command Word 4	A						A	
Trigger Word 1	Command Word 5	A							A
Trigger Word 2	Command Word 1		A		A				
Trigger Word 2	Command Word 2		A			A			
Trigger Word 2	Command Word 3		A				A		
Trigger Word 2	Command Word 4		A					A	
Trigger Word 2	Command Word 5		A						A
Trigger Word 3	Command Word 1			A	A				
Trigger Word 3	Command Word 2			A		A			
Trigger Word 3	Command Word 3			A			A		
Trigger Word 3	Command Word 4			A				A	
Trigger Word 3	Command Word 5			A					A

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

SINGLE-TRIGGER WORD-SPOTTING (SWS) MODE

When configured for Single-Trigger Word-Spotting mode, the Voice Direct™ II will, upon power-up or reset, check to see if any words have been trained. If so, then Voice Direct™ II will immediately start listening (in word-spotting mode) for the trigger word. If no words have been trained, then the Voice Direct™ II will wait for the user to train one or more words. After training, the -RECOG pin can be brought low to initiate Word-Spotting recognition.

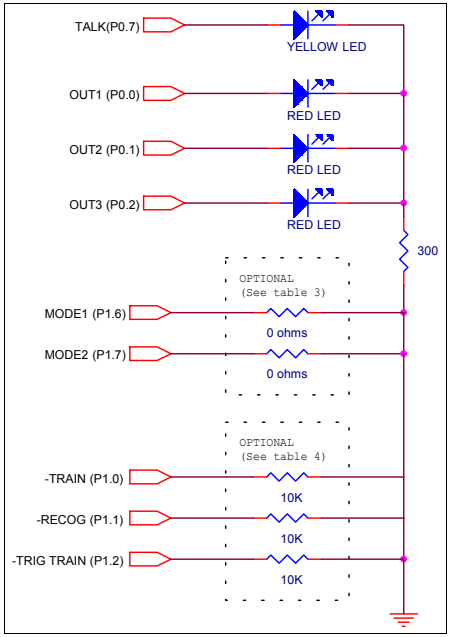
It is important to note that the module is only listening when the TALK LED is lit. Use the LED as an indicator of when you can speak the trigger and command words, especially when first using the module. After successful recognition of the trigger word, there is a 3 second window during which Voice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

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

Recognition Word	OUT1	OUT2	OUT3	OUT4	OUT5	OUT6	OUT7	OUT8
Trigger + Command Word 01 (or Trigger + No Command Words Trained)	A							
Trigger + Command Word 02		A						
Trigger + Command Word 03			A					
Trigger + Command Word 04				A				
Trigger + Command Word 05					A			
Trigger + Command Word 06						A		
Trigger + Command Word 07							A	
Trigger + Command Word 08								A
Trigger + Command Word 09	A							A
Trigger + Command Word 10		A						A
Trigger + Command Word 11			A					A
Trigger + Command Word 12				A				A
Trigger + Command Word 13					A			A
Trigger + Command Word 14						A		A
Trigger + Command Word 15							A	A

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

Figure 4 – Basic Application Schematic



MODE SELECTION

There are 4 main operating modes for the Voice Direct™ II module that can be selected depending on how voice recognition will interact with your application. These are Edge-Triggered Single Recognition (ESR) mode, Single-Trigger Continuous Listening (SCL) mode, Multi-Trigger Continuous Listening (MCL) mode, and Single-Trigger Word-Spotting (SWS) mode. The MODE1 and MODE2 pins are used to select the desired operating mode. The operating mode is set when the Voice Direct™ II is first powered up (or reset) according to the following table:

Table 3 – Voice Direct™ II Mode Configuration Pin Configuration	Operating Mode
MODE1 open circuit MODE2 open circuit	Edge-Triggered Single Recognition (ESR) mode
MODE1 pulled low MODE2 open circuit	Single-Trigger Continuous Listening (SCL) mode
MODE1 open circuit MODE2 pulled low	Multi-Trigger Continuous Listening (MCL) mode
MODE1 pulled low MODE2 pulled low	Single-Trigger Word-Spotting (SWS) mode

Operating instructions can be found in sections 4, 5 and 6 of this manual.

TRAINING AND RECOGNITION SELECTIVITY

The –TRAIN, -TRIG TRAIN and –RECOG pins control the selectivity and sensitivity of the Voice Direct™ II. The configuration parameters are set when the Voice Direct™ II is first powered up (or reset) according to the following table:

Table 2 – Voice Direct™ II Motherboard Pin Assignments

Connector	Pin #	Name	Description	I/O
J1	01	+3.3V	3.3 Volt (+) power supply connection	I
J1	02	+3.3V	3.3 Volt (+) power supply connection	I
J1	03	OUT1(P0.0)	Output 1 or 9 (Active high)	O
J1	04	-TRAIN(P1.0)	Start SD Training / Configure Training/Recognition Sensitivity (see below)	I
J1	05	OUT2(P0.1)	Output 2 or 10 (Active high)	O
J1	06	-RECOG(P1.1)	Start Recognition / Configure Microphone Sensitivity (see below)	I
J1	07	OUT3(P0.2)	Output 3 or 11 (Active high)	O
J1	08	-TRIG TRAIN(P1.2)	Start Trigger Word Training (SCL, MCL, SWS modes only) / Configure CL Performance (see below)	I
J1	09	OUT4(P0.3)	Output 4 or 12 (Active high)	O
J1	10	OUT6(P1.3)	Output 6 or 14 (Active high)	O
J1	11	OUT5(P0.4)	Output 5 or 13 (Active high)	O
J1	12	OUT7(P1.4)	Output 7 or 15 (Active high)	O
J1	13	-	Not Used – Leave Open	-
J1	14	OUT8(P1.5)	Output 8 (Active high)	O
J1	15	-	Not Used – Leave Open	-
J1	16	MODE1(P1.6)	Mode Selection Pin 1	I
J1	17	TALK(P0.7)	Output Active High When Voice Direct™ II is Listening For User Speech	O
J1	18	MODE2(P1.7)	Mode Selection Pin 2	I
J1	19	GND	Ground connection	I
J1	20	GND	Ground connection	I
J5	01	DACOUT	Analog output (unbuffered)	O
J5	02	-	Not Used – Leave Open	-
J5	03	-	Not Used – Leave Open	-
J5	04	-	Not Used – Leave Open	-
J6	01	GND	Ground connection	I
J6	02	GND	Ground connection	I
J6	03	+3.3V	3.3 Volt (+) power supply connection	I
J6	04	OUT1(P0.0)	Output 1 or 9 (Active high)	O
J6	05	-RESET	Soft Reset of Voice Direct™ II (Active low)	I
J6	06	-	Not Used – Leave Open	-
J6	07	OUT1(P0.0)	Output 1 or 9 (Active high)	O
J6	08	+3.3V	3.3 Volt (+) power supply connection	I
J6	09	GND	Ground connection	I
J6	10	GND	Ground connection	I

9. Erasing the Templates

When both the –TRAIN and –RECOG pin are pulled to GND for at least 100mS, Voice Direct™ II will erase all trained templates. Voice Direct™ II will say "Memory erased", and then exit the erase mode. Individual words or sets cannot be erased separately.

10. 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

The optimal set consists of:	Avoid sets like:	Aim for sets like:
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

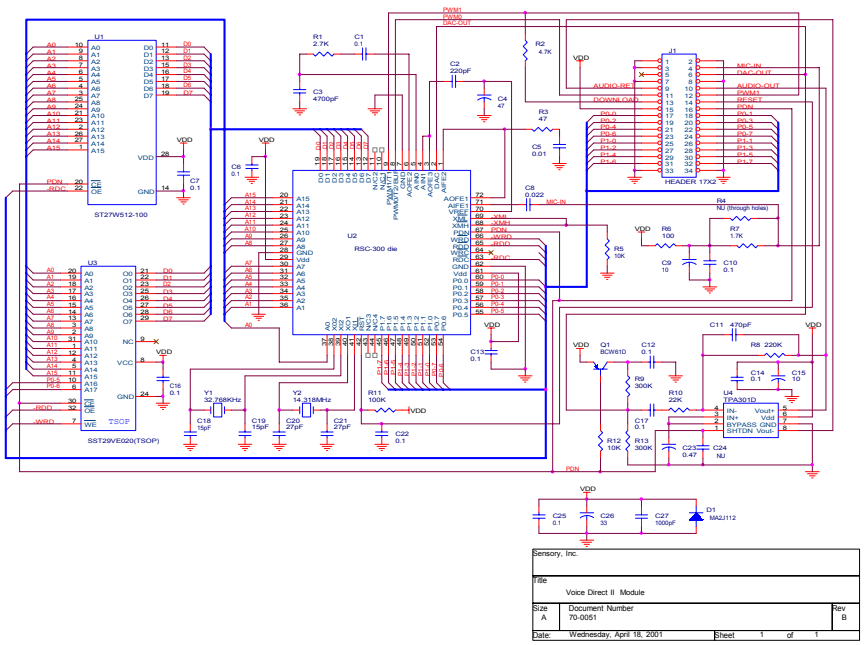
Key considerations for successful voice recording:

- 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, so the voice recording microphone should be held accordingly.
b) **Natural Voices.** Subjects should speak in a normal voice and be discouraged from sounding unusual 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, to collect voice recordings to activate an exercise machine, it would be strongly recommended to record people who sound 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. For example, a product intended for use inside a moving vehicle would benefit from samplings made inside a car with the motor running. 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.

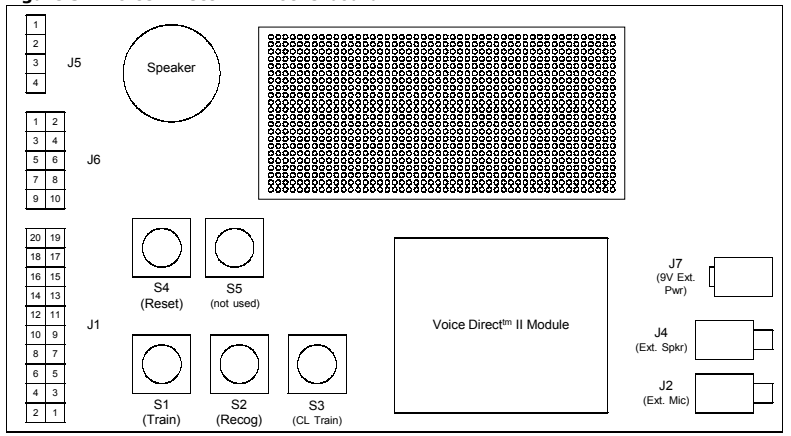
11. Schematics for Voice Direct ™ II Module and Motherboard

(An electronic version of this manual with full size schematics can be downloaded from http://www.sensoryinc.com/html/support/documentation.html)



17	OUT1(P0.0)	Output 1 or 9 (Active high)	O
18	OUT2(P0.1)	Output 2 or 10 (Active high)	O
19	OUT3(P0.2)	Output 3 or 11 (Active high)	O
20	OUT4(P0.3)	Output 4 or 12 (Active high)	O
21	OUT5(P0.4)	Output 5 or 13 (Active high)	O
22	-	Not Used – Leave Open	-
23	-	Not Used – Leave Open	-
24	TALK(P0.7)	Output Active High When Voice Direct™ II is Listening For User Speech	O
25	-TRAIN(P1.0)	Start SD Training / Configure Training/Recognition Sensitivity (see below)	I
26	-RECOG(P1.1)	Start Recognition / Configure Microphone Sensitivity (see below)	I
27	-TRIG TRAIN(P1.2)	Start Trigger Word Training (SCL, MCL, SWS modes only) / Configure CL Performance (see below)	I
28	OUT6(P1.3)	Output 6 or 14 (Active high)	O
29	OUT7(P1.4)	Output 7 or 15 (Active high)	O
30	OUT8(P1.5)	Output 8 (Active high)	O
31	MODE1(P1.6)	Mode Selection Pin 1	I
32	MODE2(P1.7)	Mode Selection Pin 2	I
33	GND	Ground connection	I
34	GND	Ground connection	I

Figure 3 – Voice Direct™ II Motherboard



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, as this 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 very thin, no more than 0.7 mm if possible.

- THIRD: The microphone should be acoustically isolated from the housing in some way. This can be accomplished by surrounding the microphone element with a spongy material such as rubber or foam. Mounting it 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.

6. Motherboard and Module Pinout and Characteristics

Figure 2 – Voice Direct™ II Module

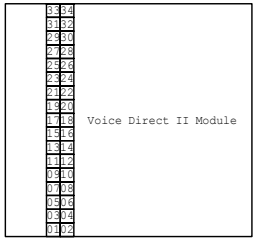
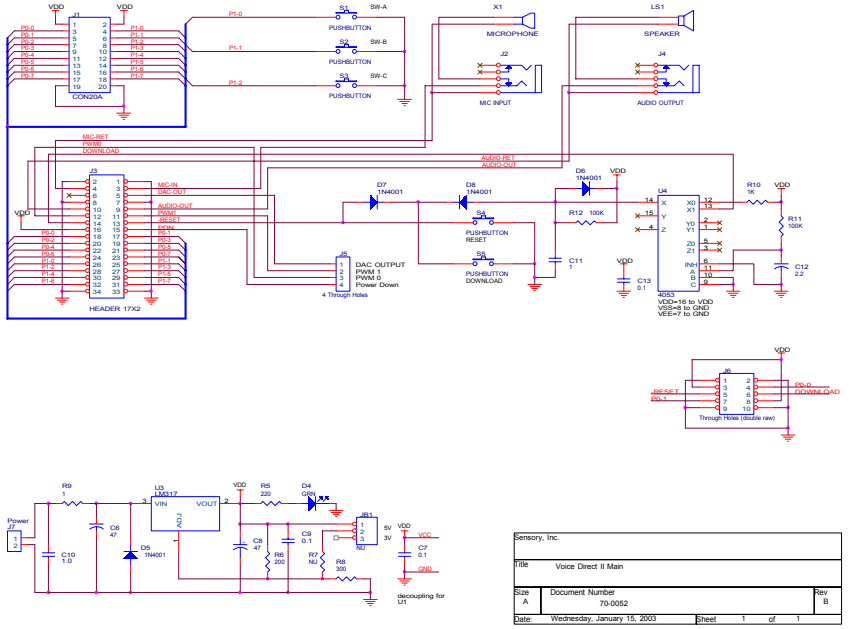


Table 1 – Voice Direct™ II Module Pin Assignments

Pin #	Name	Description	I/O
01	GND	Ground connection	I
02	GND	Ground connection	I
03	MIC RETURN	Analog Ground. For noise reasons, analog and digital grounds should connect together only at the Voice Direct™ II	-
04	MIC IN	Microphone Input Connection	I
05	-	Not Used – Leave Open	-
06	DACOUT	Analog output (unbuffered)	O
07	GND	Ground connection	I
08	GND	Ground connection	I
09	SPEAKER (-)	Speaker connection (negative)	O
10	SPEAKER (+)	Speaker connection (positive)	O
11	-	Not Used – Leave Open	-
12	-	Not Used – Leave Open	-
13	-	Not Used – Leave Open	-
14	-RESET	Soft Reset of Voice Direct™ II (Active low)	I
15	+3.3V	3.3 Volt (+) power supply connection	I
16	-	Not Used – Leave Open	-



12. Disclaimers

WARNING
This kit is intended for use by consumers with experience in electronic design. 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 part purpose. Customer must rely on customer's own skill or judgment in selecting suitable products for customer. To the extent warranty is applicable, such warranty shall be limited to 90 days from the date of purchase.

13. Voice Direct™ II Speech Recognition Kit Limited Warranty

The **Voice Direct™ II** 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 become obsolete or is no longer in production and deemed irreparable, 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: **1991 Russell Ave., Santa Clara, CA 95054**. The Buyer will pay all shipping and other charges or assessments for the return of the Product to Sensory, Inc.

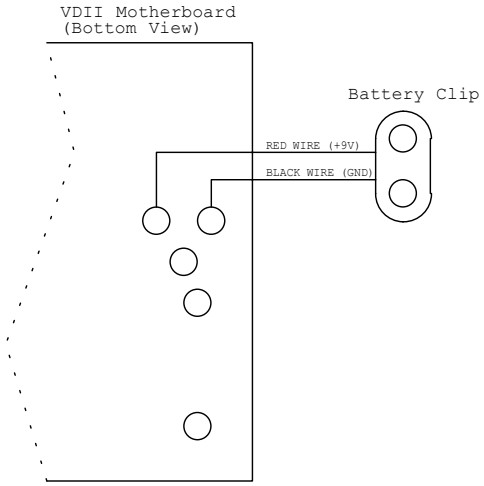
Limitation of Warranty
The foregoing warranty shall not apply to defects resulting from maintenance performed by anyone other than Sensory, Inc., modifications made by Buyer or any third party, Buyer supplied software or interfacing, misuse, abuse, accident, mishandling, operation outside the environmental 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™ II 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.

Figure 1 – Installing The Included Battery Clip



In addition to the included battery clip, the user has the option of using a standard 9V wall adapter (not included). The wall adapter can be plugged into J7 directly eliminating the need for using the battery clip.

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. For more information, refer to the DC Characteristics in Section 6.
- Keep in mind that 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™ II 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™ II.
- With proper product construction, Voice Direct™ II 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.

Steps 1-3 above are repeated for each word during training. Voice Direct™ II stores the average of two training patterns for each word to improve accuracy. Before a new template is stored, 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™ II, like all 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™ II 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™ II minimizes substitution errors but may produce more rejection errors. When the training selectivity level is set to its highest value, Voice Direct™ II 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.

5. Building a Basic Application

CONSTRUCTION

Take into consideration the fixture options for the Voice Direct™ II motherboard and/or module. You may mount the motherboard and module together, or detach the module and install it by itself. Using the module alone may increase the additional parts required for operation. Other factors will also include the size of your application, and how compact and transportable it will need to be.

- One possible method of construction is to use a small project box, mounting the speaker, microphone, pushbuttons, and motherboard and/or 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 perforated or prototyping 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 pushbuttons, 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

The Voice Direct™ II module (the small board that is plugged into the motherboard) is designed to run at 3.3V. Two 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.

The Voice Direct™ II motherboard (the large board that the module plugs into) has a 3.3V regulator installed on it, so it can be powered by the included 9V battery clip. Refer to figure 1 for instructions on soldering the included 9V battery clip to the Voice Direct™ II motherboard.

DANGER!: IF YOU UNPLUG THE VOICE DIRECT™ II MODULE AND RUN IT INDEPENDENTLY FROM THE MOTHERBOARD, DO NOT ATTEMPT TO POWER IT ABOVE 3.3V OR IT WILL BE DAMAGED!

14. SENSORY Software End User License Agreement

Important: this software end user license agreement ("EULA") is a legal agreement between you and Sensory. Read it carefully before completing the installation process and using the software. It provides a license to use the software and contains warranty information and liability disclaimers. By installing and using the software, you are confirming your acceptance of the software and agreeing to become bound by the terms of this agreement. If you do not agree to be bound by these terms, then select the "cancel" button and do not install the software.

1. Definitions

(a) "Sensory" means Sensory, Inc. and its suppliers and licensors, if any.
(b) "Not For Resale (NFR) Version" means a version of the Software, so identified, to be used to review and evaluate the Software, only.
(c) "Software" means the software program supplied by Sensory herewith, which may also include documentation, associated media, printed materials, and online and electronic documentation.

2. License

This EULA allows you to:

(a) Install and use the Software on a single computer; OR install and store the Software on a storage device, such as a network server, used only to run or install the Software on your other computers over an internal network, provided you have a license for each separate computer on which the Software is installed or run from the storage device. A license for the Software may not be shared or used concurrently on different computers.
(b) Make one copy of the Software in machine-readable form solely for backup purposes. You must reproduce on any such copy all copyright notices and any other proprietary legends on the original copy of the Software.

3. License Restrictions

(a) Other than as set forth in Section 2, you may not make or distribute copies of the Software, or electronically transfer the Software from one computer to another or over a network.
(b) You may not decompile, reverse engineer, disassemble, or otherwise reduce the Software to a human-perceivable form.
(c) You may not sell, rent, lease, transfer or sublicense the Software.
(d) You may not modify the Software or create derivative works based upon the Software.
(e) You may not export the Software into any country prohibited by the United States Export Administration Act and the regulations there under
(f) In the event that you fail to comply with this EULA, Sensory may terminate the license and you must destroy all copies of the Software.

4. Upgrades

If this copy of the Software is an upgrade from an earlier version of the Software, it is provided to you on a license exchange basis. You agree by your installation and use of this copy of the Software to voluntarily terminate your earlier EULA and that you will not continue to use the earlier version of the Software or transfer it to another person or entity.

5. Ownership

The foregoing license gives you limited rights to use the Software. Sensory and its suppliers retain all right, title and interest, including all copyrights, in and to the Software and all copies thereof. All rights not specifically granted in this EULA, including Federal and International Copyrights, are reserved by Sensory and its suppliers.

6. Limited warranty and disclaimer

(a) Limited warranty. Sensory warrants that, for a period of ninety (90) days from the date of delivery (as evidenced by a copy of your receipt): (i) when used with a recommended hardware configuration, the software will perform in substantial conformance with the documentation supplied with the software; and (ii) that the physical media on which the software is furnished will be free from defects in materials and workmanship under normal use.
(b) No other warranty, except as set forth in the foregoing limited warranty, sensory and its suppliers disclaim all other warranties, either express or implied, or otherwise including the warranties of merchantability and fitness for a particular purpose. Also, there is no warranty of noninfringement, title or quiet enjoyment. If applicable law implies any warranties with respect to the software, all such warranties are limited in duration to ninety (90) days from the date of delivery. No oral or written information or advice given by sensory, its dealers, distributors, agents or employees shall create a warranty or in any way increase the scope of this warranty.
(c) Some states (USA only) do not allow the exclusion of implied warranties, so the above exclusion may not apply to you. This warranty gives you specific legal rights and you may also have other legal rights that vary from state to state.

7. Exclusive Remedy

Your exclusive remedy under Section 6 is to return the Software to the place you acquired it, with a copy of your receipt and a description of the problem. Sensory will use reasonable commercial efforts to supply you with a replacement copy of the Software that substantially conforms to the documentation, provide a replacement for defective media, or refund to you your purchase price for the Software, at its option. Sensory shall have no responsibility if the Software has been altered in any way, if the media has been damaged by accident, abuse or misapplication, or if the failure arises out of use of the Software with other than a recommended hardware configuration.

8. Limitation of liability.

(a) Neither Sensory nor its suppliers shall be liable to you or any third party for any indirect, special, incidental or consequential damages (including damages for loss of business, loss of profits, business, interruption or the like), arising out of the use or inability to use the software or this EULA based on any theory of liability including breach of contract, breach of warranty, tort (including negligence), product liability or otherwise, even if sensory or its representatives have been advised of the possibility of such damages and even if a remedy set forth herein is found to have failed of its essential purpose.

(a) Sensory's total liability to you for actual damages for any cause whatsoever will be limited to the greater of $500 us dollars or the amount paid by you for the software that caused such damage.

(b) (USA only) some states do not allow the limitation or exclusion of liability for incidental of consequential damages, so the above limitation or exclusion may not apply to you and you may also have other legal rights that vary from state to state.

9. Basis of Bargain

The Limited Warranty, Exclusive Remedies and Limited Liability set forth above are fundamental elements of the basis of the agreement between Sensory and you. Sensory would not be able to provide the Software on an economic basis without such limitations.

10. U.S. GOVERNMENT RESTRICTED RIGHTS LEGEND

This Software and the documentation are provided with "RESTRICTED RIGHTS". Use, duplication, or disclosure by the U.S. Government is subject to restrictions as set forth in this EULA and as provided in DFARS 227.7202-1(a) and 227.7202-3(a) (1995), DFARS 252.227-7013 (c)(1)(ii)(OCT 1988), FAR 12.212(a)(1995), FAR 52.227-19, or FAR 52.227-14, as applicable. Manufacturer: Sensory, Inc., 1991 Russell Ave., Santa Clara, CA 95054.

11. Consumer End Users Only (outside of the USA)

The limitations or exclusions of warranties and liability contained in this EULA do not affect or prejudice the statutory rights of a consumer, i.e., a person acquiring goods otherwise than in the course of a business.

12. General Provisions

This EULA shall be governed by the internal laws of the State of California. This EULA contains the complete agreement between the parties with respect to the subject matter hereof, and supersedes all prior or contemporaneous agreements or understandings, whether oral or written. All questions concerning this EULA shall be directed to: **Sensory, Inc., 1991 Russell Ave., Santa Clara, CA 95054.** Attention: **General Counsel**.

1. Introduction

Thank you for purchasing the Voice Direct™ II Speech Recognition Kit. We are sure that you will appreciate how easily basic voice recognition capabilities can be added to your electronic project. For more in-depth development, we suggest you upgrade to the Voice Extreme™ Toolkit, which offers additional speech technologies and a stand-alone microcontroller.

This Quick Setup Guide is intended to provide a basic reference for the capabilities of the Voice Direct™ II SRK, and will help start you on the way to utilizing voice activation quickly and easily. For more detailed technical information about the Voice Direct™ II module and Sensory's award winning speech recognition technologies, please visit our web site at ***www.sensoryinc.com.***

2. Features

- Edge-Triggered, Continuous Listening and Word-Spotting speech recognition technologies
- Minimal external components
- Recognizes up to 15 words or phrases (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
- Quick setup time

3. Contents

One Voice Direct™ II Motherboard
One Voice Direct™ II Module
One Battery Clip
Three 10K resistors
One 470 ohm resistor
Three Red LEDs
One Yellow LED
One Quickstart Manual

4. About Speech Recognition

Voice Direct™ II 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™ II is based on a digital reconstruction of the voice command. Each word to be recognized must first be *trained.* During training, Voice Direct™ II constructs a *template* representing the individual speaker's unique sound pattern for each specific word or phrase to be recognized. Templates are stored in non-volatile flash memory. During recognition a new template is produced and compared to the stored templates to determine which word was spoken. Voice Direct™ II features integrated speech prompting for both training and recognition operations, allowing the development of sophisticated interactive products with minimal programming. Voice Direct™ II performs the following operations when recognizing a word:

1. The audio signal (spoken word or phrase) is externally amplified and filtered and then supplied to the analog inputs of the Voice Direct™ II, which converts the analog waveforms to digital samples.
2. Voice Direct™ II analyzes the speech signal samples and generates a template of information representing significant speech elements.
3. Voice Direct™ II increases or decreases the gain of the external amplifier as needed to maintain signal quality.
4. The unknown template is compared to each of the trained templates to determine the one template that provides the best match.
5. If the best match template gives a score above a pre-defined threshold, Voice Direct™ II chooses the word associated with that template. If no template provides a match above threshold, a special "no match" value is chosen.

15. 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 and tools, SC series of speech microcontrollers, plus a line of easy-to-implement chips that can be pin-configured or controlled by an external host microcontroller. Sensory's software technologies run on a variety of microcontrollers and DSPs.

RSC Microcontrollers and Tools

The RSC product line contains low-cost 8-bit speech-optimized microcontrollers designed for use in consumer electronics. All members of the RSC family are fully integrated and include A/D, pre-amplifier, D/A, ROM, and RAM circuitry. The RSC family can perform a full range of speech/audio functions including speech recognition, speaker verification, speech and music synthesis, and voice record/playback. The family is supported by a complete suite of evaluation tools and development kits.

SC Microcontrollers and Tools

The **SC-6x** product line features the highest quality speech synthesis ICs at the lowest data rate in the industry. The line includes a 12.32 MIPS processor for high-quality low data-rate speech compression and MIDI music synthesis, with plenty of power left over for other processor and control functions. Members of the SC-6x line can store as much as 37 minutes of speech on chip and include as much as 64 I/O pins for external interfacing. Integrating this broad range of features onto a single chip enables developers to create products with high quality, long duration speech at very competitive price points.

Application Specific Standard Products (ASSPs)

- ***Voice Direct™ II*** provides inexpensive speaker-dependent speech recognition and speech synthesis. This easy-to-use, pin-configurable chip requires no custom programming and can recognize up to 15 words in 4 pin-programmable modes. Ideal for speaker-dependent command and control of household consumer products, Voice Direct™ II is part of a complete product line that includes the IC, module, and Voice Direct™ II Speech Recognition Kit.
- ***Voice Extreme™*** simplifies the creation of fully custom speech-enabled products by offering developers the capability of programming the chip in a high-level C-like language. Program code, speech data, and even record and playback information can be stored on a single off-chip Flash memory. Based on Sensory's RSC-364 speech processor, Voice Extreme includes a highly efficient on-chip code interpreter, and is supported by a comprehensive suite of low-cost development tools.

Software and Technology

- ***Voice Activation™*** micro footprint software provides advanced speech technology on a variety of microcontroller and DSP platforms. A flexible design with a broad range of technologies allows manufacturers to easily integrate speech functionality into consumer electronic products.
- ***Fluent Speech™*** small footprint software recognizes up to 50,000 words; offers Animated Speech with the ability to automate enunciation and articulation; performs text-to-speech synthesis in either male or female voices; provides noise and echo cancellation, performs Wordspotting for natural language usage; offers telephone barge-in; and provides continuous digit recognition.

Important notices

Reasonable efforts have been made to verify the accuracy of information contained herein, however no guarantee can be made of accuracy or applicability. Sensory reserves the right to change any specification or description contained herein.

Voice Direct™ II
Speech Recognition Kit

Quick Start Manual

SENSORY®

P/N 80-0240-A