

## I. Introduction

In order to guarantee a robust speech recognition product, careful testing is needed. This document describes the recommended procedures for testing speech recognition designs incorporating the Sensory RSC-4128 and RSC-464 microprocessors.

## II. Test Materials

The following items are required for speech recognition testing:

- The product to be tested,
- A controlled testing environment, free from unwanted outside noise, preferably a soundproofed room,
- Pre-recorded “Golden” voice .WAV files of the words to be recognized to provide repeatable results. The Golden Voices should be:
  - Recorded at CD quality, (44100 Hz, 16 bit, Mono),
  - Recorded in a studio setting to be free from unwanted background noise and echoes,
  - Recorded at a level between 50 to 75% of maximum amplitude. Recordings must not be too loud as to saturate the microphone input, nor too soft as to lack bits of resolution,
  - Recorded without any special effects such as reverb, preemphasis, chorus or loudness. These may make the recordings “sound” better, but they are not real-world and the RSC4128 speech recognizer cannot recognize them as well as unprocessed, natural voices.
  - Recorded from first language speakers of the same country, age, gender and accents as the target market of the product (speaker-independent only).
- Prerecorded or generated noise .WAV files (optional). The following types of noise are commonly used for testing:
  - White noise,
  - Pink noise,
  - Music (classical, pop and/or rap),
  - Recorded conversations or television.
- A THX certified speaker for guaranteed flat frequency response of Golden voices and noise,
- A mixer to combine Golden voices and noise in separate channels. The mixed must be able to independently set the output level of each channel,
- A sound level meter, capable of measuring dbA and dbC, and averaging over 1 at least second.

## III. Test Setup

Set up the test as follows:

- Place the product to be tested on a table in the middle of the test room,
- Place the THX certified speaker about 2 meters away. Orient the speaker so it points in the general direction, but does not point directly at the product to be tested. (The exact distance between the speaker and product is not critical since the sound level is measured at the product)
- Place the sound level meter on top of, or next to, the product to be tested.

- All sound and noise level measurements should be averaged over as long a period as possible. Most sound level meters can be configured to average over one second. In the case where a noise source has long periods of high amplitude followed by long periods of low amplitude (such as music or recorded conversations), it may be useful to select a representative 5-10 second sample and play it back in an endless loop.

#### IV. Test Guidelines

- Before doing any testing, determine the purpose of testing and what the acceptable failure rate should be. No speech recognition system is perfect - even human recognition stops working in the presence of sufficient noise.
- Typically, there are three general categories of tests, each of which has its own specific purpose and parameters. See Table 1.

Test	Design Verification Test	Production Test	QA Audit Test
<b>Purpose</b>	Verify the SR capability of a given design	Simple pass/fail for a production unit	Check a random sample to ensure a lot meets QA standards.
<b>Test Parameters</b>	Test all words in a variety of background conditions	Test 1-2 words in low noise	Test all words in very low noise
<b>Test SNR</b>	80 dB signal 60 dB noise	70 dB signal 50 dB noise	75 dB signal 50 dB noise
<b>Expected results</b>	Allow for 5-10% failure (or as spec allows)	Must pass within 2 tries	All words must pass within 2 tries

Table 1 – Test Categories

- The Sensory RSC-4128 and RSC-464 ICs work best when the Signal-to-Noise Ratio (SNR) is 18 dB or higher. Lower SNR translates into lower accuracy. Figure 1 shows the approximate Recognition vs. SNR relationship.

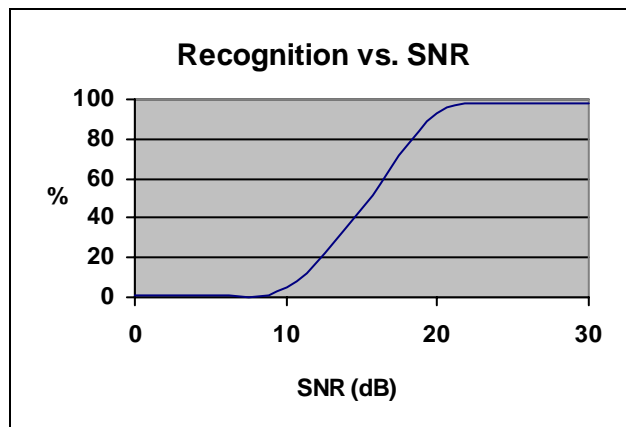


Figure 1

## 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 technologies. It is designed for cost-sensitive consumer-electronic applications such as home electronics, home automation, toys, and personal communication. The product line includes the award-winning RSC-4x general-purpose microcontrollers and tools, the *VR Stamp™* 40 pin DIP module and tools, the SC series of speech and music synthesis microcontrollers. Our suite of software development kits are designed to run on non-Sensory processors and DSP's, and support most popular operating systems.

### RSC Microcontrollers and Tools

The RSC product family 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 recording/playback. The family is supported by a complete suite of evaluation and development toolkits.

### Speech Recognition Modules and Tools

The VR Stamp™ is a complete speech recognition module based on the RSC-4x and is ideal for fast design and easy production. A low-noise audio channel and standardized 40-pin DIP footprint allow rapid prototyping, less debugging, and shorter time to market. The *VR Stamp Toolkit* includes everything needed to get started today, including VR Stamps, Module Programming Board, sample applications, and a complete set of development tools featuring the Phyton IDE and limited-life C compiler, QuickSynthesis™ 4 and Quick T2SI-Lite™ speech tools.

### SC Microcontrollers and Tools

The SC-6x product family 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 processing and control functions. Members of the SC-6x line can store as much as 37 minutes of speech on-chip and include as many as 64 I/O pins for external interfacing. Integrating this broad range of features into a single chip enables developers to create products with high quality, long duration speech at very competitive price points.

### FluentSoft™ Technology

FluentSoft™ Recognizer is the engine powering the FluentSoft™ SDK. It provides a noise-robust, large-vocabulary, speaker-independent solution with continuous digit recognition and word-spotting capabilities. This small-footprint software recognizes up to 5,000 words; runs on non-Sensory processors including Intel XScale, TI OMAP, and ARM9 platforms; and supports operating systems such as MS Windows, Linux, and Symbian.

### 3Dmsg™ Technology

3Dmsg's ([www.3Dmsg.com](http://www.3Dmsg.com)) Animated Speech technology offers animated avatars with advanced speech recognition and synthesis capabilities for use in smartphones, language trainers, and kiosk applications. Facial expressions can be configured to show emotions and lip synchronization can be automatically driven from voice or text data.

### Important notices:

Sensory Incorporated (Sensory, Inc.) reserves the right to make changes, without notice, including circuits, standard cells, and/or software, described or contained herein in order to improve design and/or performance. Sensory, Inc. assumes no responsibility or liability for the use of any of these products, conveys no license or title under any patent, copyright, or mask work right to these products, and makes no representations or warranties that these products are free from patent, copyright, or mask work right infringement, unless otherwise specified. Applications that are described herein for any of these products are for illustrative purposes only. Sensory, Inc. makes no representation or warranty that such applications will be suitable for the specified use without further testing or modification.

### Safety Policy:

Sensory, Inc. products are not designed for use in any systems where malfunction of a Sensory, Inc. product can reasonably be expected to result in a personal injury, including but not limited to life support appliances and devices. Sensory, Inc. customers using or selling Sensory Incorporated products for use in such applications do so at their own risk and agree to fully indemnify Sensory, Inc. for any damages resulting from such improper use or sale.



575 N. Pastoria Ave., Sunnyvale, CA 94085  
Tel: (408) 625-3300 Fax: (408) 625-3350

© 2006 SENSORY, INC. ALL RIGHT RESERVED.  
Sensory is registered by the U.S. Patent and  
Trademark Office.

All other trademarks or registered trademarks are the  
property of their respective owners.