February 17, 2004

Lakshman One
School of Engineering Science
Simon Fraser University
Burnaby, British Columbia
V5A 1S6

Re: ENSC 440 Project Functional Specification –Voice Recognition System in an MP3 Player

Dear Mr. One:

The attached document is a *Functional Specification for a voice recognition system in MP3 player*. We are currently working with Start Labs Inc., whose product, an MP3 player, is to be controlled by the voice of the user. Our design is the voice recognition module of the product. We will ensure that the design is in accordance with Start Labs Inc.'s expectation and needs in the most effective ways.

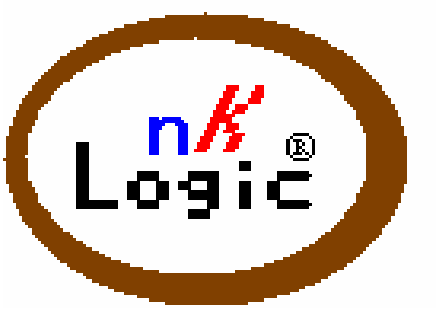
This proposal includes a system overview, a tentative budget and funding. We have found a few viable solutions or designs and they are discussed and compared in the system overview. A tentative schedule of the project progress is also added in this document

nK Logic consists of two experienced senior engineering students: Won Kang and Garet Kim. We are looking forward to hearing your feedback and suggestions. Please feel free to contact me by phone at (604) 785-5933 or by e-mail at gkim@sfu.ca. Thank you for you attention.

Sincerely,

Garet Kim
nK Logic

Enclosure: *Functional Specification for a voice recognition system in a MP3 player*




# Functional Specification for a

# Voice Recognition System in MP3 Players

**Project Team:** Won Kang
Garet Kim

**Contact Person:** Garet Kim
gkim@sfu.ca

**Submitted to:** Lakshman One – ENSC 440
Nakul Verma – ENSC 440
Mike Sjoerdsma – ENSC 305
School of Engineering Science
Simon Fraser University

**Issued date:** February 17, 2004

**Revision:** 1.1

# Executive Summary

We, nK Logic Group, are committed to building an advanced voice control unit (VCU), which is to be mounted on the MP3 Player from Start Labs Inc. After a careful examination, we have narrowed down our possible solutions, and, in order to shorten the development schedule, we have already purchased Voice Extreme™ Development Toolkit from Sensory Inc. This toolkit is divided into two main boards: Voice Extreme™ Development Board and Voice Extreme™ Module. The functionality of our VCU will be described on the assumption of using the custom IC version of this Voice Extreme™ Module.

To help understand the functionalities of the whole VCU, we hereby described a detailed analysis of what functions are required by the system in the following pages. In addition to the definition of main functionalities, we also included some other important aspects, such as power consumption, extendibility, interfaces, and limitations.

# Table of Contents

# List of Figures



# List of Tables

# 1. Introduction

The goal of this project is to build a convenient and reliable Voice Control Unit which enables users to have total control over the MP3 Player. The most important factor is the system reliability given the fact that even today's most advanced voice recognition technologies does not guarantee 100 percent hit ratio. However, considering the fact that our VCU is designed specifically for an MP3 player, the power consumption and size limitations are some other critical factors as well.

Although the actual implementation will be based on a custom IC designed for voice recognition, the basic functional requirements do not depend on the specific choice of the IC. In terms of hardware, it requires MPU, memory, IO and external interface, debugging utilities, and power, just as with other micro-controller operated devices. On the other hand, the software parts can be divided into algorithms, flow of operations, and optimized coding. Each of these components is analyzed thoroughly in this functional specification.

Unlike other projects, this project is simply to implement the required functions on a single voice recognition development board, Sensory Inc.'s Voice Extreme™. Therefore, considerations for product development such as aesthetics and physical dimensions are not taken into account here. This project is focused on implementing functions specified by Start Lab Inc. and testing them to ensure the reliability of the system.

## 1.1 Scope

This functional specification document lists the requirement of the Voice Control Unit for an MP3 player. It includes the hardware and software requirement, physical limitations, operating condition, and testing plans.

## 1.2 Intended Audience

This functional specification document is intended for use by the design engineers of nK Logic Group. It is also an agreement between the Start Labs Inc. and nK Logic Group for the previously proposed Voice Recognition System.

# 2. System Overview

The following diagram shows the general configuration for the hardware of the VCU. During the design process, the connections and compositions can be altered, but the essential functionality of each component will remain the same.

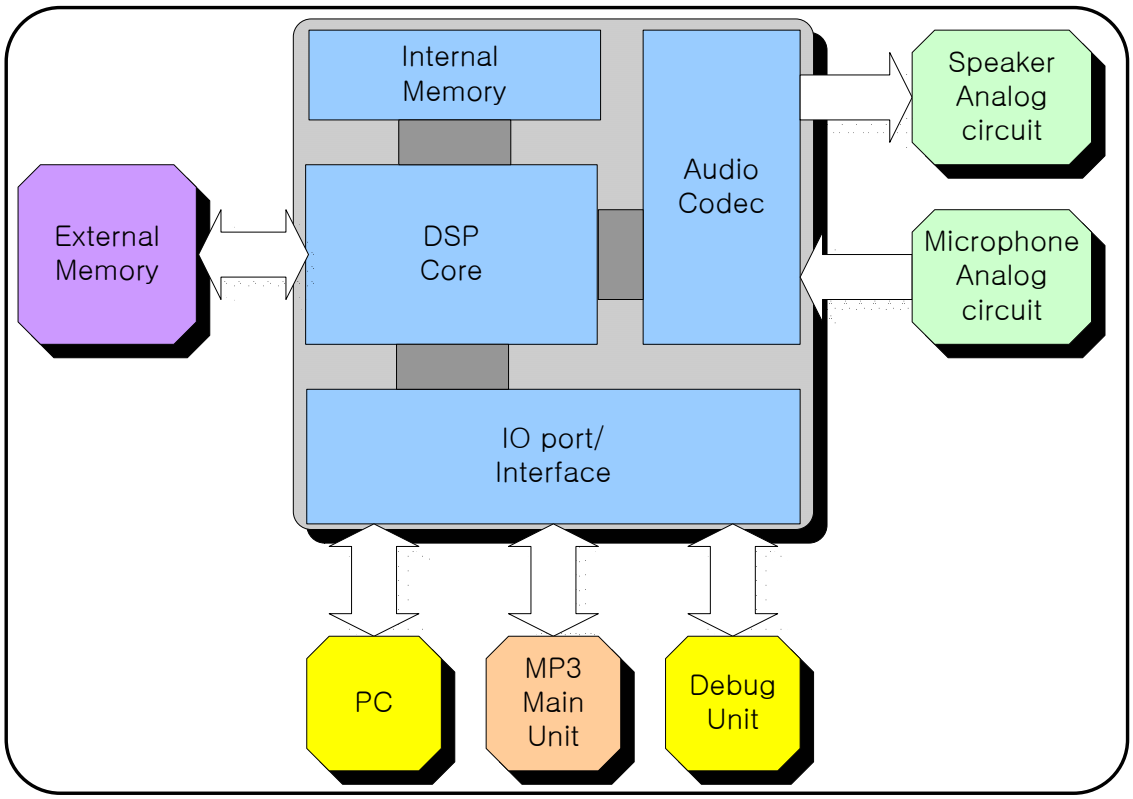



**Figure 1 : General Configuration**

In addition, some of the hardware components can be replaced by the ones that already exist on the MP3 Main Unit, resulting in the efficient usage of resources.

On the basis of this hardware block, necessary software will be composed on a PC base environment and downloaded to the hardware. Since a voice recognition program is not a simple sequential program, it is better to have a high level language interpreter. Initially the VCU will have only Speaker Dependent (SD) features, but it can be extended to have full Speech Technologies such as Speaker Independent (SI), Speaker Verification (SV), and Word Spot (WS).

# 3. Hardware Specifications

Simplified voice recognition can be viewed as a series of sequential processes as shown in the figure below.

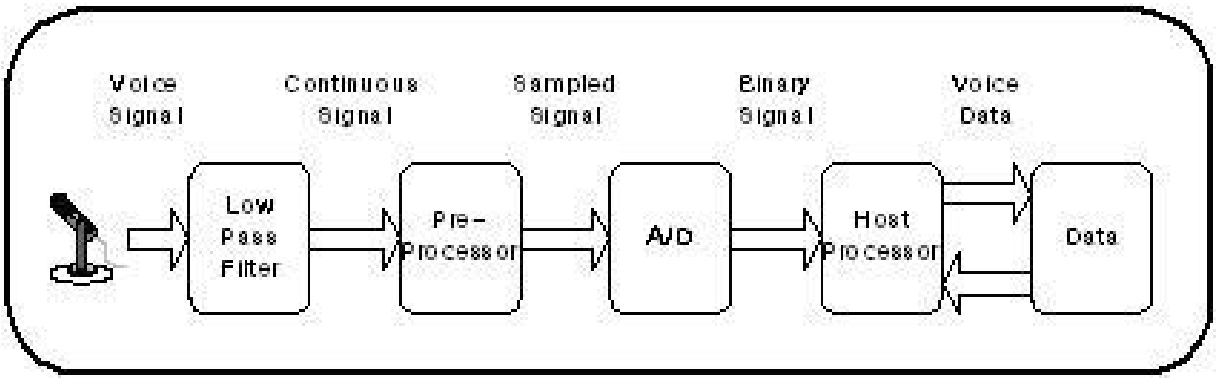



**Figure 2 : Configuration of Voice Recognition Process**

The following is the detailed functionalities of those hardware components mentioned in the overview section.

## 3.1 MPU

Micro-processor is the heart of the hardware for the VCU. It integrates all other sub-systems and decides the actions that the VCU performs. It has high level language interpreter to understand the voice recognition programs, to operate the various IO and peripherals, and to execute multiple programs. Enough processing power to perform the actions in real time is essential for the MPU, and low power consumption is another key requirement.

## 3.2 Memory

Memory can be divided into RAM and ROM. RAM is the memory space required for the MPU to process the software programs, and ROM is the memory space for the storage of source code and voice data.

Type and size of RAM depend on the specific IC (DSP) chosen, but typically it is integrated into the IC so that no external RAM is required.

The size of the ROM depends on the number of available commands rather than the code size since the voice weights occupy much more space. Flash ROM is appropriate for the development stage due to the repeated usage, but One Time Programmable (OTP) type is good enough for the actual production. Other alternative is to share a portion of the Flash from MP3 Player.

However, the maximum size of the ROM is limited by the number of address lines of the chosen MPU.

## 3.3 Audio Codec

Audio Codec is an indispensable requirement for every digital audio device. It converts analog voice signal into digital codes so that the CPU can perform diverse actions. Also it converts digital signal into the analog audio signal so that the speaker can play the sound. It does the linear PCM conversion, but advanced systems require PCMU or PCMA conversions as well. Due to the frequent usages, it is often integrated into the ASIC for the convenience of the developers.

## 3.4 Speaker

Speaker is the front end of the system to get the user input. The use of speaker is to generate announcement message or user prompt. The message can be in the form of short sentences, or a number of BEEP sound. It can be replaced by the MP3 player's speaker.

## 3.5 Microphone

Microphone is the front end of the system to get the user input. It is connected to the A/D converter (Audio Codec) to provide word data. For the reliable performance of the system, it is important to have maximum Signal to Noise Ratio (SNR). The quality of the microphone, PCB layout, and gain settings affect the SNR significantly, and additional bandpass filter must be implemented to increase noise immunity.

## 3.6 Interface / IO port

The VCU requires a number of interfaces to interact with other devices. The following table is a brief summary of each.

**Table 1: Required Interface**

| Interface | Bit | Function |
|---|---|---|
| SPI | 3 | Serial Downloading of source code<br>Connection with real time debugging terminal |
| Input | 1<br>1 | Toggle the activation of VCU<br>System reset |
| Output | 3 | Output LEDs for simple result |
| TBD | TBD | Interface with MP3 player |
| Reserved | 6 | Future enhancement |

Some of these interfaces and IO ports are not needed in the production stage.

# 4. Software Specifications

By using the custom ICs, the role of the software is to take care of the flow of the operation. The main disadvantage of using custom IC is the fact that the access to the core technologies is not permitted. Therefore, it is necessary to optimize the given parameters to the maximum.

## 4.1 Algorithm for Voice Recognition

Various algorithms and methodologies have been invented and many ASICs are available for the applications.

## 4.2 Command Grammar

For a successful implementation of voice recognition system, it is important to have a well-defined syntax for grammar representation. Following table illustrates possible command sets to be implemented on the system. Further consultation with a linguistics specialist may be needed to improve grammar and sentence representation.

**Table 2: Command Sets**

| **1. Playback** | |
|---|---|
| | |
| Play | |
| "Play" | Play the song on the top of the playlist (after power-off), or resume |
| "Play (1, 2…)" | Play n-th song on the playlist |
| | |
| Pause | |
| "Pause" | Pause the current playing song |
| | |
| Stop | |
| "Stop" | Stop the current playing song. |
| | |
| Rewind | |
| "Back (seconds)" | Rewind by certain seconds |
| | |
| Fast Forward | |
| "Forward (seconds)" | Fast Forward by certain seconds |
| | |
| Previous | |
| "Previous (1, 2, …)" | Play previous tracks |
| | |
| Next | |
| "Next (1, 2…)" | Play next tracks |
| | |
| | |
| | |

| | |
|---|---|
| **2. Volume** | |
| | |
| "Volume Up" | Volume up by 1 level, un-mute. |
| "Volume Down" | Volume down by 1 level |
| "Volume (0,1,…..,9,10)" | Volume scaled from 0 to 10.<br>User can change the volume to a larger extent. |
| "Mute" | Mute the player.<br>To un-mute, say "Volume Up" |
| | |
| | |
| **3. Option** | |
| | |
| "Normal" | Play normally. (i.e. No repeat, no shuffle) |
| "Repeat" | Repeat the currently playing song |
| "Repeat All" | Repeat the entire playlist |
| "Repeat Off" | Escape from repeat mode |
| "Shuffle (or Random)" | Play songs randomly |
| "Shuffle (or Random) Off" | Escape from shuffle mode |
| | |
| | |
| **4. Equalizer** | Change the equalizer's settings |
| | |
| "Mode Classic" | Set the equalizer to classic mode |
| "Mode Latin" | Set the equalizer to Latin mode |
| "Mode Rock" | Set the equalizer to Rock mode |
| "Mode Dance" | Set the equalizer to Dance mode |
| "Mode Club" | Set the equalizer to Club mode |
| "Mode Full Base" | Set the equalizer to maximize the base |
| "Mode Full Treble" | Set the equalizer to maximize the treble |
| "Mode Large Hall" | Set the equalizer to Large Hall mode |
| "Mode Live" | Set the equalizer to Live mode |
| "Mode Party" | Set the equalizer to Party mode |
| "Mode Reggae" | Set the equalizer to Reggae mode |
| "Mode Soft" | Set the equalizer to Soft mode |

| | |
|---|---|
| "Mode Techno" | Set the equalizer to Techno mode |
| "Mode Pop" | Set the equalizer to Pop mode |
| | |
| | |
| **5. Power** | |
| | |
| "Power Off" | Power off |
| "Reset (or Reboot)" | Reboot the software |
| | |
| | |
| **6. Skin** | |
| | |
| "Skin (1, 2, …)" | Change the background of the LCD display |

## 4.3 Memory Usage

Along with CPU processing power, memory usage is a vital characteristic to describe a voice recognition system. 2MB of flash memory is expected to be enough for the current command set, but accurate measurement of the memory usage is necessary for optimal design.

## 4.4 Debugger

In addition to the LEDs and the speaker, it is better to have a real time debugger to analyze the system. There must be a fast, real-time, pc-based debugger to allow the developers to examine the operations of units.

# 5. Other System Specifications

## 5.1 Safety

In developing an audio device, keeping the amplified audio within allowable range is most critical. In order to prevent any damage to the eardrum, the system should never output any unwanted loud sound.

## 5.1 Others

Since nK Logic develops a prototype voice recognition module for Start Lab Inc., size, weight, and packaging of the prototype module are not of nK Logic's concern. System performance, operation conditions, and other facts about the system are the same as the development board's specifications provided by the manufacturer. The following is a summary of those specifications.

**Table 3: Specifications of Voice Extreme™**

| | |
|---|---|
| **Manufacturer** | Sensory Inc. |
| **Product** | Voice Extreme™ |
| **Core** | 8-bit CPU |
| **Add. Memory Reqd** | 2MB Flash |
| **Maskable ROM** | 64KB |
| **Internal ROM** | N/A |
| **Speech Duration (Max.)** | 100 sec. (ext. flash) |
| **RAM** | 2.5 KB |
| **I/O** | 14 |
| **Key Technologies** | SI, SD, SV, CL |
| **SI words on chip** | 350 (ext. flash) |
| **SD words on chip** | 1900 (ext. flash) |
| **Packages** | TQFP-64 |
| **100k die price** | <$1.50 |
| **Power dissipation** | 3.0V, 10mA |
| **Notes** | ASSP of RSC-3X |
| **Operating Temperature** | 0 to 70 degrees C |

# 6. Test Plan

## 6.1 Hardware Testing

For this project, the hardware will be the evaluation board obtained directly from Sensory Inc. The only available testing is to adjust the microphone gain settings.

a) Replace the microphone resistors with variable gain resistor
b) Measure the hit ratio for comparison
c) Repeat the same for both indoor and outdoor

## 6.2 Software Testing

### 6.2.1 Functionality

a) Test every combination of possible commands for its correct operation.
b) Repeat the same for both indoor and outdoor.

### 6.2.2 Reliability

a) Test each for 3 times with at least 5 different speakers.
b) Measure the error rate for the commands
c) Adjust the command and weight
d) Re-test those with high error rates.

# 7. Conclusion

The functional specifications in this document outlined expectations from both nK Logic and Start Lab Inc. in terms of the functionality of the system.

Start Lab Inc. can find out specifications of the development board, Voice Extreme$^{TM}$ by Sensory Inc., in this document. We also listed commands that the user can use to prompt the MP3 player.

To ensure high success ratio in voice recognition, many linguistic considerations need to be put when choosing commands. nK Logic is planning on consulting with a linguistics professor regarding this issue.

# Glossary

| | |
|---|---|
| Voice Extreme™ | Sensory Inc.'s application specific IC that enables intuitive programming of interactive speech application |
| Weight | Samples of voice data required for speech synthesis and pattern comparison |
| Hit Ratio | Ratio that a voice recognition unit successfully recognizes the user's commands. |

# Acronyms

| | |
|---|---|
| MP3 | Moving Picture Experts Group Layer-3 Audio (audio file format/extension) |
| WS | Word Spot |
| OTP | One Time Programmable |
| ASIC | Application Specific Integrated Circuit |
| PCM | Pulse-Code Modulation |
| PCB | Printed Circuit Board |
| SNR | Signal to Noise Ratio |
| LED | Light Emitting Diode |
| CL | Continuous Listening |
| SD | Speaker dependent |
| SI | Speaker Independent |
| SV | Speaker Verification |