SEES PROJECT

**Spatial Echolocation Enhancement System**

**Software Design Document**

University of Victoria

CENG/ELEC/SENG 399 Fall 2014

Design Team 14

Daniel Faulkner - V00778450 - SENG

Paulo Tabarro - V00810880 - CENG

Jason Lim - V00785426 - SENG

Adalberto Outeiro - V00810849 - ELEC

Rajpal Chauhan - V00762290 - ELEC

# Introduction

This document will describe the software architecture of the Spatial Echolocation Enhancement System. This software is intended to provide a logical back end for the SEES device that will process input data from the SEES headset peripheral and provide appropriate audio feedback to the user accordingly. The topics covered in this document include descriptions of the input data sources the software will draw from as well as the outputs it is expected to provide. Also included are descriptions of the software’s main operating modes as well as key audio processing algorithms that will be employed by the software.

# Software System Inputs

The SEES software takes in 3 primary inputs: A depth image stream; a Head Related Transfer Function (HRTF) profile; and a set of possible user control inputs. These inputs have been summarized in Table 1. The depth image stream is provided by the SEES headset peripheral at 30 frames per second and is 640x480 pixels in size. Each pixel on the image contains 16 bits representing depth. Using the points on the depth image, the software builds a model of objects and obstacles relative to the user’s head. The model is then used to choose transfer functions from the HRTF profile stored on the user’s mobile device. Finally, the software computes spatialized audio cues based on the object model using the chosen transfer functions. In addition to the sensor input and HRTF profile, the software also takes in inputs from the user to allow actions such as switching operating modes or loading and changing HRTF profiles.

**Table 1:** System Inputs

|  |  |
| --- | --- |
|  | **Description** |
| **Image Stream** | A 640x480, 16bpp depth image stream at 30 fps. |
| **HRTF Profile** | A Head Related Transfer Function profile containing transfer functions necessary for computing binaural audio about the user`s head. |
| **User Input** | Control inputs from the user to change operating modes or configure the software. |

# Software System Outputs

With the exception of a user interface that may include graphical, aural, and haptic feedback, the SEES software’s primary output is one or more binaural audio cues transmitted to the user as a single stereo audio stream through a pair of headphones. Human hearing has a maximum frequency range below 20000Hz, so in order to maximize the effectiveness of audio cues and minimize aliasing, audio is sampled at 44000Hz [1].

**Table 2:** System Outputs

|  |  |
| --- | --- |
|  | **Description** |
| **Audio Stream** | A binaural audio signal at 44000Hz |
| **Menu Feedback** | System response to user input in the form of graphical, aural, and/or haptic feedback. |

# Operating Modes

The SEES software has 3 primary operating modes outlined in Table 3. The software can be modeled as a state machine, as shown in Figure 1. The entry point of the software when launched is in the idle state, where the system is not actively interfacing with any of the inputs or outputs. The user can transition into the sensing or the configuration modes from here.

tain binaural audio software and associated control software (~6 MB)

Note that in the following paragraphs term ‘selection’ may imply pressing a button on the graphical interface or prompting the system using a voice command.

Configuration mode is entered from idle mode by the user selecting the *Configure* option. The primary feature of configuration mode is to allow the user to set up his or her HRTF profile. This can be accomplished in one of two ways:

* Download the profile from a database of pre-recorded HRTF profiles.
* Import a custom profile from storage on the mobile device

Configuration mode can be exited by the user selecting the *Back* option. Any changes will be saved. The user will be returned to idle mode.

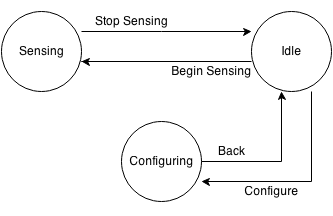
Sensing mode is entered from idle mode by the user selecting the *Begin Sensing* option. When in sensing mode, the system will actively interface with the input and output. The system will poll the sensor input and perform convolutions in real time to transform the input into binaural audio with reverberation. The system will send this to the output. Sensing mode can be exited by the user selecting the *Stop Sensing* option, which will return the software to idle mode.

**Table 3:** Main Running States

|  |  |
| --- | --- |
| **State** | **Description** |
| **Sensing** | System will interface with sensors and headphones, performing convolutions to transform the visual signal to an auditory signal based on the user’s HRTF profile. System will monitor user input to switch to idle. |
| **Idle** | System will monitor user input to switch to sensing mode or to enter configuration. |
| **Configuration** | System will allow user to set up operating modes - specifically, user can download, import an HRTF profile. |

**Table 4:** State Transitions

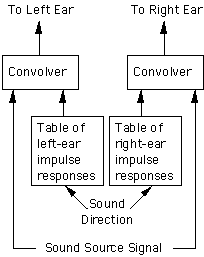
|  |  |
| --- | --- |
| **Transition** | **Description** |
| **Sensing → Idle** | System will transition from **Sensing** to **Idle** when user selects the *Stop Sensing* option in the application. |
| **Idle → Sensing** | System will transition from **Idle** to **Sensing** when user selects the *Begin Sensing* option in the application. |
| **Idle → Configuration** | System will transition from **Idle** to **Configuration** when user selects the *Configure* option in the application. |
| **Configuration → Idle** | System will transition from **Configuration** to **Idle** when user selects the *Back* option in the application. |



**Figure 1:** State Machine diagram for SEES software operating modes

# Audio Processing

The SEES Software uses binaural audio filtering to generate spatialized audio cues when listened to through a pair of headphones. The process involves taking a single mono audio cue used as a base and filtering it with two transfer functions for the left and right ears to obtain two audio signals that are then streamed out to headphones as a stereo audio. Selecting the two transfer functions to use in filtering the audio is dependent on the desired perceived location of the spatialized audio cue. Typical HRTF profiles contain transfer functions for each of the subject’s ears recorded at various azimuths and elevations around the subject’s head. Thus, given a pair of horizontal coordinates in which an spatialized audio cue is desired, the corresponding transfer function can be chosen appropriately. Figure 2 summarizes the process of generating binaural audio by showing how the process can be implemented in a system known as a convolvotron:



**Figure 2:** A binaural audio convolvotron as illustrated by the

University of Calgary CIPIC Interface Laboratory HRTF Database [2]

As traditional convolution is CPU intensive to calculate, convolution for the left and right audio signals will be performed using fourier analysis via the convolution theorem. Using the convolution theorem, convolution between two signals can instead be calculated as a point-wise multiplication of the two signal’s fourier representations [3]. This changes the number of point multiplications required in the calculation from n2 down to n; however, it also introduces the need for efficiently calculating the fourier transform of the streaming audio signal.

Assuming that the HRTF fourier representations can be pre-calculated, then producing spatialized audio requires 1 forward transform for the mono input channel and 2 backward transforms for the stereo output channels. Therefore, in order to produce continuous filtered audio for a signal provided at 44000Hz it is necessary to calculate fourier transforms of 132000 data points per second. To this end, the software will use the native C FFTW fourier transform subroutine library. The FFTW library has been benchmarked as being able to perform nearly 4000 4096-point fourier transforms per second on a 1.06 GHz processor core [4]. This can equate to roughly 16000000 points per second on a mobile phone processor of the same strength making the library an ideal choice for implementing a continuous convolvotron on a mobile platform.

# 

# 

# Conclusion

The SEES software serves both as an audio processing unit and as a logical back end to the SEES system. It takes sensor data from the headset, an HRTF profile, and user commands as inputs and provides spatialized audio cues as output. For control, the software exposes a user interface with optional graphical, aural, or tactile cues to facilitate accessible control of the system. Using this interface, user’s can switch the software between idle and sensing modes as desired. An additional configuration allows users to adjust the behaviour of the system and select the HRTF profile used for generating spatialized audio cues. Actual convolution between the HRTF profile data and the audio signal are performed via the FFTW fourier transform library to allow rapid audio filtering of multiple audio cues simultaneously.

Appendix A - References

[1] H. E. Heffner and R. S. Heffner, “Hearing ranges of laboratory animals,” in *Journal of the*

*American Association for Laboratory Animal Science*. American Association for

Laboratory Animal Science, 2007, vol. 46, num. 1, 2007, pp. 20-22.

[2] [Weisstein, Eric W.](http://mathworld.wolfram.com/about/author.html) "Convolution Theorem." From [MathWorld](http://mathworld.wolfram.com/)--A Wolfram Web Resource.

<http://mathworld.wolfram.com/ConvolutionTheorem.html>

[3] Department of Electrical and Computer Engineering, University of California Davis. (2011,

Feb. 25). *HRTF-Based Systems* [Online]. Available:

<http://interface.idav.ucdavis.edu/sound/tutorial/hrtfsys.html>

[4] M. Frigo and S. G. Johnson. (2014, Mar. 04). *1.06 GHz PowerPC 7447A, gcc-3.4* [Online].

Available: <http://www.fftw.org/speed/G4-1.06GHz-gcc3.4/>

# Appendix B - List of Tables

|  |  |  |
| --- | --- | --- |
| ***Figure*** | ***Description*** | ***Page Number*** |
| **Table 1** | System Inputs | 2 |
| **Table 2** | System Outputs | 2 |
| **Table 3** | Main Running States | 3 |
| **Table 4** | State Transitions | 3 |

# Appendix C - List of Figures

|  |  |  |
| --- | --- | --- |
| ***Figure*** | ***Description*** | ***Page Number*** |
| **Figure 1** | State Machine diagram for SEES software operating modes | 4 |
| **Figure 2** | A binaural audio convolvotron as illustrated by the  University of Calgary CIPIC Interface Laboratory HRTF Database | 5 |