

## Enhanced Speech Recognition Evaluation Program (E-SREP)

Richard L. Lynch, Andrew L. Kun

### Introduction

The Project54<sup>1</sup> Speech Recognition Evaluation Program (SREP)<sup>2</sup> was developed by Brett Vinciguerra to predict which speech recognition engine and microphone combination would recognize Project54 speech commands with the most accuracy and efficiency.



Figure 1 - Original SREP Set Up

In its original form, SREP had high resource requirements. Ideally, it required a PC with two sound cards, two sets of speakers (i.e. 4 total), and a microphone. The microphone and speakers had to be set up in the cabin of a police cruiser for accurate measurements. An example set up is shown in Figure 1.

SREP functioned by playing car noises (e.g. noise from the A/C, open windows, radio, engine) through one set of speakers, and spoken Project54 commands through the other set of speakers. The speakers and microphone were installed in police cruisers to accurately simulate the acoustics of a police cruiser. The Project54 code would then relay phrases recognized by the speech recognition engine back to SREP.

To ensure accurate results, the placement of the speakers and the volume of each sound card/speaker combination should remain fixed over all tests. However, some of these tests might

be performed well into the future – comparing future engines to present day engines. Ensuring identical testing equipment and testing conditions over long periods of time is not feasible. To alleviate this problem, a calibration module was created. The calibration module adjusts each sound card/speaker combination to a specific volume level relative to the microphone. This was the first objective of E-SREP.

The second objective of E-SREP was to reduce the resource requirements of SREP – spare police cruisers are not very common, and they are difficult to move from office to office. Additionally, PCs with multiple sound cards are not very common. E-SREP eliminates the sound cards and police cruiser from the resource requirements.

## **Background**

Scientific experiments typically involve three kinds of variables – independent, dependent, and controlled. The independent variable is the variable that is intentionally manipulated. The dependent variable is the variable being observed, and typically is a function of the independent variable. Control variables are variables that are held constant since they can affect the dependent variable. With respect to SREP, the independent variable was the speech recognition engine, the dependent variable was accuracy, and the control variables included the testing environment and the aggregate sound card/speaker volume. In order to ensure valid results, the control variables must be held constant over all tests. In its original form, SREP had no means of ensuring this.

A simple solution to this problem would be to set both sound cards at their maximum volume, superglue the speaker volume controls in place, and always use the same sound cards/speakers for SREP testing. This solution may produce reasonable results, assuming the original sound cards and speakers are always available for testing, and the characteristics of the sound cards and speakers do not change with age. However, a more elegant solution would be to calibrate the sound card/speaker volumes to a specific volume level with each test.

## Sound Card/Speaker Volume Calibration

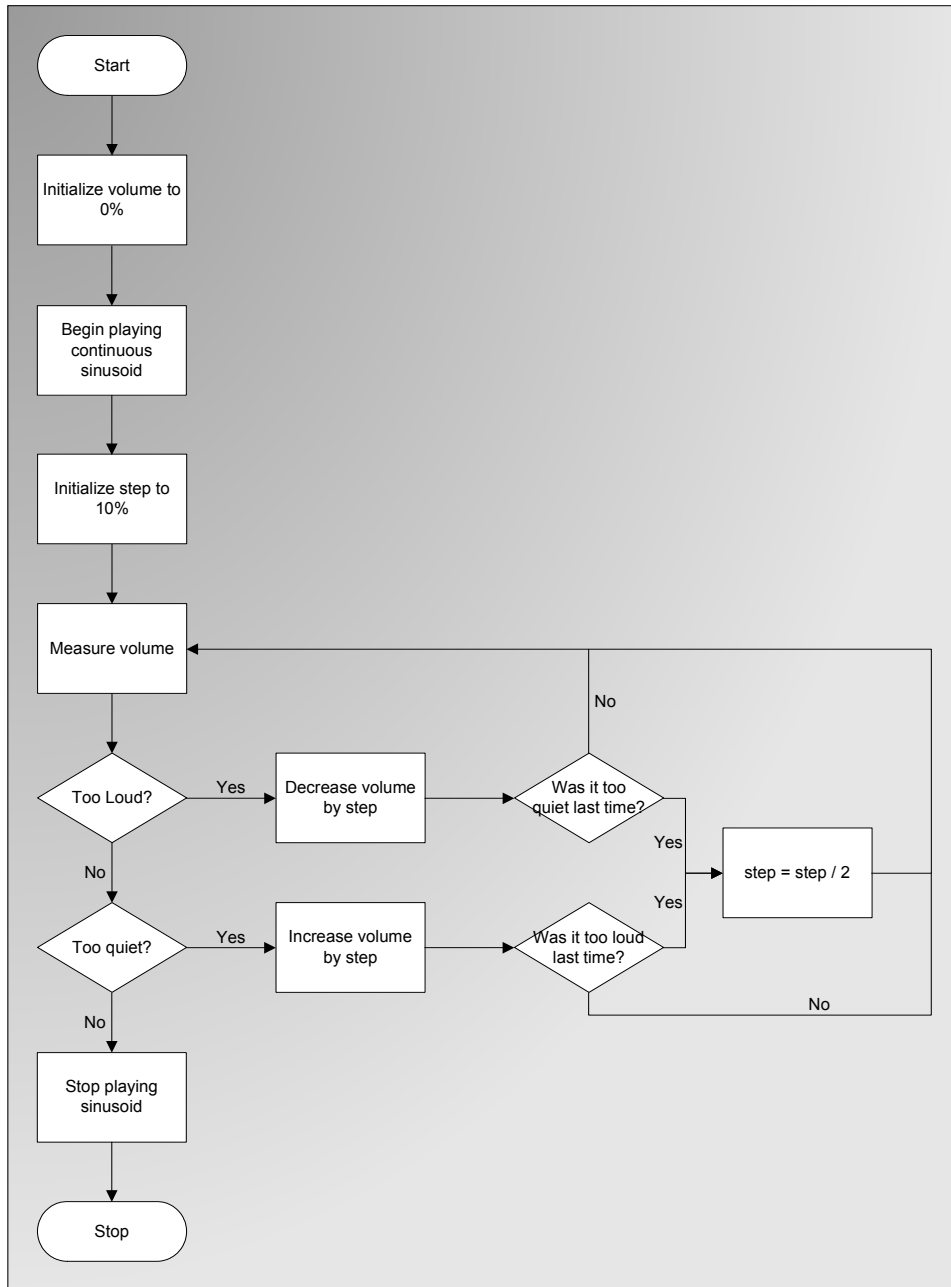
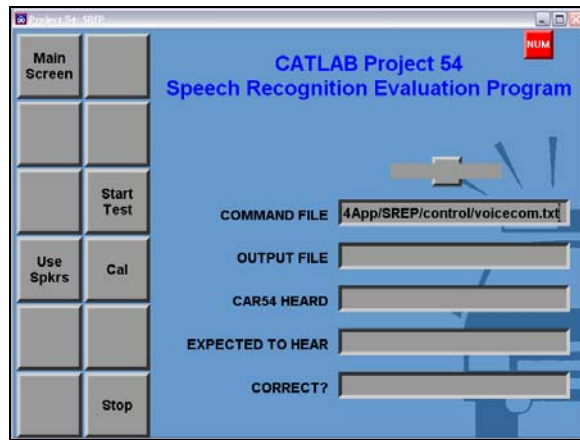


Figure 2 - Volume Calibration Flowchart

E-SREP calibrates a sound card/speaker combination by playing a sinusoid through the speakers, measuring the volume level at the microphone, and adjusting the master volume on the sound card until the volume level is within an adjustable tolerance of the target volume (Figure 2). E-SREP assumes that any automatic gain control mechanisms on the microphone or sound card have been disabled.



**Figure 3 - E-SREP Window**

Pressing the “Cal” button in the E-SREP application triggers the calibration function (Figure 3). The “Test in Progress” control will light up until the calibration is complete.

### **Eliminating the Sound Cards, Speakers, and Police Cruisers**

Volume calibration is one step forward in reducing the variability of the control variables, but to truly ensure the environment remains the same over all of the tests, it is necessary to eliminate the real world part of the environment altogether. This is accomplished by directly feeding premixed audio (noise + speech) into the speech recognition engines. The premixed audio was prerecorded by the Project54 code under actual usage conditions.

To achieve this, a special “FEEDFILE” message was added to the SpeechIO Project54 application. The “FEEDFILE” message would instruct the SpeechIO application to feed the specified WAV file to the speech recognition engine. Recognized commands would then be routed to E-SREP by the normal Project54 message routing mechanisms.

To activate premixed mode, the “Use Spkrs” toggle button should be popped out (gray, instead of white). Premixed mode configuration is identical to regular SREP configuration, except the command file should specify premixed audio files in place of the speech files. E-SREP will ignore any noise files in premixed mode.

## **Conclusion and Future Work**

E-SREP has successfully improved the quality of SREP's results by improving the control variables. It can either calibrate the volume of the sound cards and speakers, or it can eliminate the control variables altogether by eliminating the need for the sound cards and speakers.

Volume calibration takes less than a minute.

In the future, E-SREP would benefit from a faster calibration algorithm. Additionally, more work is needed to ensure that each speech recognition engine receives a similar amount of training.

## **References**

<sup>1</sup> Kun, Andrew L, Miller, W. Thomas, and Lenharth, William H, "Project54: Introducing advanced technologies in the police cruiser," IEEE Spring VTC2002, Birmingham, AL, May 6-9, 2002

<sup>2</sup> Vinciguerra, Brett, "A comparison of commercial speech recognition components for use with the Project54 system," MS Thesis, University of New Hampshire, Durham, 2002