A Speech Processing Research Platform for Android Based Smart Phones and Tablets

Roger Chappel¹, Kuldip Paliwal²

Signal Processing Laboratory, Griffith University
¹roger.chappel@griffithuni.edu.au, ²k.paliwal@griffith.edu.au

Abstract

This paper presents a new research and education platform for speech processing. The platform is called Speech Enhancement for Android (SEA) and incorporates past and present speech enhancement techniques applied to recorded speech corrupted by real world noise sources. Researchers, students and teaching staff can use this platform to perform speech enhancement and observe its effects on the linguistic content of speech. This paper outlines the speech processing strategies implemented in SEA along with discussing the benefits of this new interactive method to present speech processing theory and research. Index Terms: speech enhancement, speech processing, spectrogram, Android, smart phone

1. Introduction

Speech Enhancement for Android (SEA) is an interactive education and research platform for speech processing. Where the user can load or record input speech, corrupt it by real world noise sources and enhance it with one of the five included speech enhancement methods, all on the users personal Android based device. At each stage in the process, a spectrogram of the speech is created and displayed to the user, after which the user can play the corresponding audio. SEA can be configured by everyone one from a first year undergraduate student to a final year post-graduate student at varying levels of theory comprehension. Each processing step allows complete abstraction to the user, enabling the user to adjust each individual parameter of the spectrogram, speech enhancement and noise estimation. These features allow more advanced methods and parameters to be decomposed into a simple output. Not only helping students to gain an intuitive interpretation of speech enhancement, but it enables researchers in the field of speech enhancement to examine configuration changes rapidly, saving them time on programming and interpretation. All results can be saved and shared via e-mail or any other supported file transfer method, providing the ability to compare multiple methods and treatments.

Recent advances in low-powered CPU design have introduced a new era of high performance hand held devices, capable of intensive graphics rendering with on board graphics processing units (GPU) and complex operations using multi-threaded technology [1, 2]. This is quickly opening doors to allow researchers to develop portable tools to test audio processing work in the field. SEA has been developed in the Java-based Android operating system because in 2011, Android smart phones accounted for 52.5 % of the global smart phone market share, double that of 2010 [3]. Additionally, SEA also supports a wide range of Android based tablets as they begin to gain momentum in the global market [4].

This generation of students attending universities have become accustomed to a high level of technology in their pockets. So integrating content into preferred communication technologies which appeal to them through visual and interactive means may also influence personal intrinsic learning which could accelerate students understanding and comprehension of content [5, 6].

The remainder of this paper is organised as follows. Section 2 describes the speech processing methods implemented in SEA. Section 3 introduces the user interface (UI) and interactive components with screen shots. Section 4 describes how SEA can be used as a useful research tool. Section 5 takes a look into the features that are under development for future versions of SEA. And Section 6 makes some concluding comments.

2. Speech Processing Methods

The speech processing methods implemented in SEA are composed of two subsection, they are speech analysis and speech modification. For speech analysis purposes, SEA provides some valuable features. These are

- The spectrogram of speech and
- The short-time analysis of speech.

Once speech is loaded, the spectrogram of speech is generated and displayed. This spectrogram is generated with a 32 ms frame length, 4 ms frame shift and a Hamming analysis window, providing an appropriate resolution to display the phonetic and linguistic fine-structure of speech. The user can switch to the frame analysis mode, where they can observe the short-time Fourier magnitude spectrum associated with specific frames by dragging a cursor across the spectrogram (shown in Figure 1).

The speech modification component of SEA allows the user to corrupt recorded speech with pre-packaged, real-world noise sources at the desired level, including single and multi speaker babble noise, perform speech enhancement and observe the effects on the linguistic content in the resulting speech. Additionally, the user can simply enhance speech recorded in a noisy environment. The speech enhancement methods included in the current version of SEA are

- Spectral subtraction [7],
- Spectral Wiener estimator (SW) [8],
- Minimum Mean Square Error (MMSE) short-time spectral amplitude (STSA) estimator [9],
- LogMMSE STSA estimator [10],
- MMSE + speech processing uncertainty (SPU) STSA estimator [9] and
Each with their advantages and disadvantages depending on the level and statistical properties of the noise signal. Figure 2 shows speech corrupted by white noise and enhanced using the SW estimator. The configurable nature of SEA allows the user to adjust parameters and rapidly observe its effects on a speech signal. For example, the speech enhancement methods require a noise estimate in order to estimate a set of clean discrete Fourier transform (DFT) coefficients, this noise estimate is created using a voice activity detector (VAD), whereby the noise estimate is updated in frames of speech absents [12]. The VAD has specific parameters which can be adjusted such as the update rate and noise threshold. Upon making adjustments the user can enhance and within seconds observe the resulting spectrogram and listen to the resulting speech. In addition the short-time analysis-modification-synthesis (AMS) settings can be changed. The adjustable AMS settings include

- Frame duration,
- Frame shift,
- Analysis window and
- Synthesis window.

The analysis and synthesis windows available in this version of SEA include the Rectangular, Hamming, Hanning and Blackman windows. SEA allows the user to observe the role of the Fourier magnitude and phase spectrum in respect to the impact on linguistic information. The role of the magnitude and phase spectrum for speech intelligibly has been the topic of many research papers [13, 14, 15]. In these studies it has been shown that the magnitude spectrum is known to contribute significantly more to the intelligibility of speech then the phase spectrum when a $20 \sim 40 \text{ ms}$ Hamming window is employed for speech analysis. However, the phase spectrum has been shown to have a comparable contribution to the intelligibility of speech when a low dynamic range synthesis window, i.e. rectangular window, with $\approx 1 \text{ s}$ duration is employed for speech analysis [16]. SEA allows the user to observe these findings in a quick intuitive manner by simply changing the AMS configuration and modifying the speech with one of two modification techniques. These techniques are unit magnitude and random phase modification of input speech. These techniques can be grouped into two categories which are: phase only reconstruction and magnitude only reconstruction. By randomising the phase spectrum, it can be considered as magnitude only reconstruction, whereby the original magnitude is preserved and the phase is modified. Similarly, by setting the magnitude spectrum to unity it can be considered as phase only reconstruction, whereby the original phase is preserved and the magnitude is modified. Figure 3 demonstrates magnitude only reconstruction at two different AMS frame durations, allowing the user to observe that at long window durations the magnitude spectrum no longer contains intelligible linguistic information.

### 3. Interactive Components

Every aspect of SEA has been developed for interactivity, the objective of this design method is to enable the user to interact with the speech processing methods without hindrance between speech acquisition, parameter adjustments and execution. This
is an attempt to remove the feel of SEA being a piece of software which requires the user to follow set procedures to reach the end result. SEA is designed to let the user perform their desired functions in the most logical and intuitive way. For example, once the user has recorded speech they can simply play the speech by touching the spectrogram and pressing play, or the user might want to corrupt the spectrogram with additive noise from a specific noise source such as: airport, restaurant or street; so they can touch the spectrogram and select add noise and choose the type and level of noise required. All the parameters for the spectrogram such as dynamic range and color, all the AMS and VAD settings are adjusted by sliding the quick edit menu across from its hidden location on the left of the screen. This quick edit menu is transparent and can be left open while making rapid changes, or simply closed to slide back into its hidden position and accessed when needed. Figure 4 displays the quick edit menu overlaid on the main interface.

4. A Useful Research Tool

SEA can be used as a tool to test and compare speech enhancement methods. In the background of every stage from loading of speech, corrupting it by noise and performing enhancement, SEA stores the associated audio as 8 kHz WAV files and spectrogram images as PNG files. These files can be shared via email, bluetooth, sms or any other supported file transfer service already installed on the device. Additionally the images can be viewed and edited by any compatible applications on the device.

Two other tools that SEA provides are frame analysis discussed in Section 2 and the spectrogram comparison. In both of these tools, background images can be viewed and shared. For example, if the user was interested in viewing and examining a particular frame in the frame analysis tool on their computer, they can simply press menu > share and a list of supported sharing services will appear. Similarly, if the user has completed enhancement, not only can they share the individual spectrograms from the main screen, they can press menu > view spectrograms and a new view will appear allowing you to view and listen to the clean speech, corrupted speech and enhanced speech. They can then share a single image created in the background containing all three spectrograms labeled a), b) and c) as seen in Figure 5. This means that these pictures can be used in educational material such as academic papers, power point slides or course notes. It also provides the facility to share your audio and image files globally with other colleagues.

5. Evaluation and Future work

Table 1 shows the results of a survey (based on a 4-point Likert scale) conducted over twenty-four students in a speech processing course at the University of Texas, Dallas.

The students found SEA to be a useful and enjoyable platform to use in conjunction with theoretical content. More over it allowed them to successfully draw connections between theory and practice which has helped in the development of an intuitive understanding of content. Preliminary findings from researching staff at the University of Texas and Griffith University have...
indicated that SEA is a useful platform to rapidly make changes and view the effects. The results of these subjective tests have influenced certain changes which will be implemented in Version 2.0 set to be released in July 2012. These changes include increasing interactivity and refined control. In addition, some extra features will be included:

- Complete LPC spectrum overlay in frame analysis.
- Implementation of the Chebyshev analysis and synthesis window function.
- Implementation of additional speech enhancement methods.
- Export facilities to save current settings such as AMS and VAD settings to a text file for post analysis.

### Table 1: Results of a survey evaluating the usefulness of SEA, where sd = strongly disagree, d = disagree, a = agree and sa = strongly agree.

<table>
<thead>
<tr>
<th>Feature</th>
<th>sd</th>
<th>d</th>
<th>a</th>
<th>sa</th>
</tr>
</thead>
<tbody>
<tr>
<td>is easy to use</td>
<td>0</td>
<td>0</td>
<td>29.17%</td>
<td>70.83%</td>
</tr>
<tr>
<td>contributed to learning</td>
<td>0</td>
<td>0</td>
<td>58.33%</td>
<td>41.67%</td>
</tr>
<tr>
<td>is links theory to practice</td>
<td>0</td>
<td>0</td>
<td>41.67%</td>
<td>58.33%</td>
</tr>
<tr>
<td>motivates to learn</td>
<td>0</td>
<td>0</td>
<td>25.00%</td>
<td>75.00%</td>
</tr>
<tr>
<td>develops intuitive understanding</td>
<td>0</td>
<td>0</td>
<td>37.50%</td>
<td>62.50%</td>
</tr>
<tr>
<td>is enjoyable to use</td>
<td>0</td>
<td>0</td>
<td>04.17%</td>
<td>95.83%</td>
</tr>
<tr>
<td>mean</td>
<td>0</td>
<td>0</td>
<td>32.64%</td>
<td>67.36%</td>
</tr>
</tbody>
</table>

6. Conclusions

This paper has demonstrated a new use of Android-based smart phones and tablets for education and research purposes. The aim of speech enhancement for Android (SEA) is to help researchers, teachers, and students in the field of speech processing to conduct their work quickly. SEA is a new method for presenting and processing speech, allowing users to interact with speech in real-world environments and directly draw connections between theory and practice.

7. References