Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 57 (2014) 13–38

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

# An educational platform to demonstrate speech processing techniques on Android based smart phones and tablets

Roger Chappel *, Kuldip Paliwal

*Signal Processing Laboratory, School of Engineering, Griffith University, Nathan Campus, Brisbane QLD 4111, Australia*

Received 7 November 2012; received in revised form 29 July 2013; accepted 6 August 2013
Available online 19 August 2013

**Abstract**

This work highlights the need to adapt teaching methods in digital signal processing (DSP) on speech to suit shifts in generational learning behavior, furthermore it suggests the use of integrating theory into a practical smart phone or tablet application as a means to bridge the gap between traditional teaching styles and current learning styles. The application presented here is called "Speech Enhancement for Android (SEA)" and aims at assisting in the development of an intuitive understanding of course content by allowing students to interact with theoretical concepts through their personal device. SEA not only allows the student to interact with speech processing methods, but also enables the student to interact with their surrounding environment by recording and processing their own voice. A case study on students studying DSP for speech processing found that by using SEA as an additional learning tool enhanced their understanding and helped to motivate students to engage in course work by way of having ready access to interactive content on a hand held device. This paper describes the platform in detail acting as a road-map for education institutions, and how it can be integrated into a DSP based speech processing education framework.


*Keywords:* Digital signal processing (DSP); Electronic engineering education; Speech enhancement; Android; Smart phone

## 1. Introduction

Teaching digital signal processing (DSP) is often a requirement in the undergraduate and post graduate electronic engineering curriculum.

Students undertaking this field of study cover topics such as: Discrete Fourier transform (DFT) (Nussbaumer, 1981) and spectral estimation and filter design (Talor, 1983). At Griffith university, DSP is currently taught in three stages. Firstly, the continuous time analysis of signals and systems (Soliman and Srinath, 1990). Secondly, the introduction of discrete time signals and systems incorporating digital filtering on stationary processes (Proakis and Manolakis, 2007). And, lastly advanced topics such as spectral estimation and adaptive filtering for stochastic signals (Haykin, 1991; Kay, 1993; Hayes, 1996).

Each stage is accompanied by a series of laboratory sessions, exposing the student to programming techniques which can be useful in the industry, both on a simulation environment such as Matlab (Mitra, 2006) and on a hardware based tool such as the Texas Instruments 6713 digital signal processor starter kit (DSK) (Chassaing, 2002). The practical element is often concluded with a primary project providing a valuable opportunity for students to apply prior DSP learnings to a known real world application in areas such as speech coding for telecommunication (Crochiere, 1981) and automatic speaker verification (Rosenberg, 1976). In this paper, we propose an additional tool which we can be incorporated into the DSP syllabus to aid in demonstrating fundamental principals both during laboratories and lectures along side Matlab, DSK and relevant text books. This tool is called "Speech Enhancement

* Corresponding author. Tel.: +61 434574285.
*E-mail address:* roger.chappel@giriffithuni.edu.au (R. Chappel).
*URL:* http://www.enhancementapp.com (R. Chappel).

0167-6393/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.specom.2013.08.002

for Android" (SEA) and has been developed on the Google Android platform with the intention to bring DSP principals to the increasingly popular smart phones, which now account for 52.5% of the global smart phone market share, double that of 2010 (Gartner, 2011b). Additionally, SEA has been developed to support a wide range of Android based tablets as they begin to gain momentum in the global market (Gartner, 2011a; Milanesi, 2011). The mobility and accessibility of personal hand held devices over laptops in *K*-12 education has been assessed, resulting in an increase in student engagement (Soloway et al., 2001) and an increase in effective technical education in mathematics (Yerushalmy and Ben-Zaken, 2004).

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 and multi-core technology. NVIDIA (2010), ARM (2009). As the popularity is increasing and manufacturing methods are improving, these devices are now becoming accessible to students. This is opening many opportunities for education institutions to bring low cost interactive educational tools to every student. SEA capitalises on these rising opportunities by utilizing the computing power of modern hand-held devices which otherwise would have only been available on a computer.

The objective of this tool is to provide additional flexibility in the delivery of course content to accommodate for shifts in generational learning styles and behaviors (Caillaud and Cohen, 2000; Billings et al., 2005; Rosenbaum and Rochford, 2008; Salajan et al., 2010). In addition to shifts occurring in learning behaviors, other factors such as technology trends being closely associated with lifestyle is becoming prevalent (Wellman, 2002; Hassanlou et al., 2009; Martin et al., 2011). The majority of students in established universities will have access to a smart phone which provides a portal, integrating individuals' reality with that of their online virtual-reality; where study, work, entertainment and social life amalgamate into one unifying device (Donath and Boyd, 2004; Ellison et al., 2007; Humphreys, 2010; Song et al., 2012). Studies show (White, 1989; Price et al., 2003; O'NeilJr and Perez, 2003; Lubis et al., 2010; Price and Falcao, 2011), that adapting content to suit these trends can encourage and motivate students to engage in activities and enhance their learning potential. Moreover, integrating content into preferred communication technologies which appeal to each generation through visual and interactive means may also influence personal intrinsic learning and cognitive skills which could accelerate student's understanding and comprehension of content (Sims, 1997; Joiner et al., 2006; Kong et al., 2012; Rienties et al., 2009).

SEA focuses on speech processing methods, not only introducing the student to a real world application for DSP such as the field of speech enhancement but also demonstrating the principals of analysing a human speech signal. These principals address issues such as stationarity, spectral estimation, the role of the magnitude and phase spectrum, filtering, statistical methods for noise removal and the spectrogram representation of speech. These topics extend beyond that of DSP and can be useful in other disciplines such as linguistics, speech pathology and audiology.

Another goal of this tool is to present many sophisticated signal processing techniques through a simple user interface (UI) allowing a realizable and practical interpretation of every element. It is designed to give the user complete control of all configurable settings, allowing the abstraction of the more advanced speech enhancement techniques to a simple, intuitive example of each concept. This allows researchers in the field of speech enhancement to examine configuration changes rapidly, saving them time on programming and interpretation. Results can be saved and shared via e-mail or any other supported file transfer method.

This paper is organised as follows. A review of current mobile DSP applications and a description of the development process of SEA are shown in Section 1 and 2. The primary elements and functionality of SEA are discussed in Section 4. Section 5 describes the short-time analysis–modification–synthesis (AMS) framework which is used for the speech processing in this work. Section 6 illustrates the procedure undertaken to record and display the spectrogram of a speech signal. Section 7 demonstrates the short-time analysis function of SEA, allowing the user to view the time and frequency domain representation of each frame. Section 7 describes the linear prediction, source-filter separation method used in SEA. Section 10 illustrates the procedure undertaken to convey the role of the short-time magnitude and phase spectra on speech intelligibility. Section 11 describes the speech enhancement principals behind the implementation within SEA. Section 12 details a subjective test evaluating SEA's effectiveness as a learning tool in a learning environment. Section 16 explains future objectives and challenges faced for the SEA project. And lastly, Section 17 contains a summary with concluding comments.

## 2. Related mobile platforms

The concept of mobile educational platforms in electrical engineering and bioinfomatics has been generating a lot of interest since the evolution of the smartphone (Teng and Helps, 2010; Finkelstein et al., 2010). Prior to the smartphone revolution, web-based learning aids were the central theme on providing accessibility to course content for students at home or on their own laptop while on campus through an extensive wireless network. These applications ranged from complete online courses and laboratories (Maiti et al., 2011) to simple Java applets to accompany other course content (Gonvalves and Canesin, 2001). Several online DSP learning aids have been developed (Ko et al., 2003; Jackson et al., 2001), however, only a small minority have extended their platforms to

accommodate the uprising of the smartphone and capitalize on the increasingly low cost, accessibility and mobility (Potts et al., 2011). One of the most recent developments is the extension of the Java-DSP (J-DSP) web applet to the Android (A-JDSP) (Ranganath et al., 2012) and iOS (i-JDSP) (Liu et al., 2011) platforms. The JDSP family was designed with similar objectives to SEA, in that bringing accessible interactive content to students can enhance learning potential, develop cognitive skills and help to simplify complicated theoretical concepts through visualisation. The JDSP family offers the students access to content and tools for convolution, sampling, frequency domain analysis and filter design.

The primary difference between SEA and the JDSP family is that although they share much of the underlying theory in the background processing, SEA has been developed for those students wishing to extend their DSP knowledge beyond fundamental DSP concepts to speech processing and advanced filtering. SEA enables the student to apply DSP theory to a real-world application in speech, where theoretical concepts can be examined on a speech signal and their effects on speech quality and intelligibility can be observed. SEA provides engineers and other speech based disciplines with a platform to examine real-world trade offs and considerations for different speech processing methods, helping to bridge the gap for those students who are transitioning between undergraduate studies to the industry or higher research.

## 3. Design and development

Initially, SEA was designed to demonstrate DSP speech processing methods to students through visual and intuitively interactive means. However, the effectiveness of the UI as being intuitively interactive is a broadly subjective claim. Initial informal trails of SEA on a sample of students provided vital information to improve the UI and enhance interaction between students and theoretical content.

The software supporting the underlying framework of SEA was developed in Java using the Android software development kit (SDK). The AMS framework used in SEA was firstly developed in MATLAB and then in C on a Linux based system, both methods were comprehensively tested and compared to ensure consistency and accuracy. The Android SDK framework was also exhaustively tested against the previous two frameworks which upon validation, the UI was then designed and developed. The entire design, development and trial process was then repeated after the initial trails, where major improvements were made for the UI and functionality. SEA has now undergone several revisions since its release, and has slowly evolved into a stable, enjoyable platform.

## 4. Primary elements and functionality

SEA allows users to observe four primary elements of speech processing which enable an intuitive comprehension of content and allow users to draw connections between theory and real applications. The four elements are:

1. Frame analysis: the analysis of a speech signal both in time and frequency domains for each frame.
2. Spectrogram representation of speech: the construction of an image representing phonetic structure of speech across time and frequency.
3. Modification of speech: the role of the magnitude and phase spectrum in terms of speech quality and intelligibility.
4. Speech enhancement: enhancing speech which is corrupted by real life noise sources, e.g. train, car, jet, babble, restaurant, street, airport and white noise.

These elements are unified through an interactive UI, enabling the user to record their own speech, load some pre-packaged examples or load speech from file as an input. After which they can either; add noise of their choosing and commence enhancement, modify speech using one of two different spectral modification techniques or analyse both the short-time magnitude and time domain of the speech segment. Fig. 1 shows a block diagram demonstrating the elements of SEA, from the acquisition of input speech to displaying the spectrogram of both the input speech and enhanced or modified speech. Upon the acquisition of an input source the spectrogram of the input speech is automatically generated and displayed. Similarly,

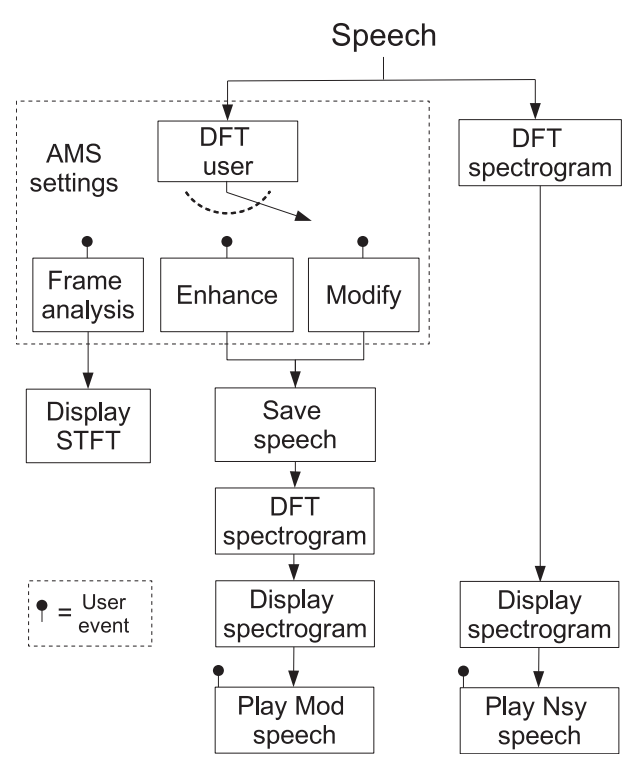


Fig. 1. Block diagram illustrating the functionality of the interactive elements within SEA.

once speech is either enhanced or modified, the resulting speech spectrogram is also automatically displayed. The modification and enhancement stage utilises the short-time Fourier AMS framework; however, the spectrogram generation and frame analysis stages only use short-time Fourier analysis, as no modification or synthesis is required.

### 4.1. Main interfaces

Fig. 2 shows three different interfaces the user is presented upon specific touch interactions. Fig. 2(a) is the first interface the user is presented upon opening SEA. There are three primary buttons each assigned to specific tasks. Starting from the bottom left, the 'Record' button allows the user to record 3 seconds of speech and upon completion, the spectrogram is generated from the recorded speech and displayed in the top view. Next is the 'Enhance' button; upon a touch event, the enhance button takes user enhancement settings selected in the quick edit menu (Fig. 2(b) and performs enhancement accordingly, producing an enhanced audio file and displaying the enhanced spectrogram. To the lower right of Fig. 2(a) is the 'Modify' button; upon a touch event, an additional options menu is presented to the user, allowing them to select one of two frequency based modification techniques; 1. random phase and 2. unit magnitude.

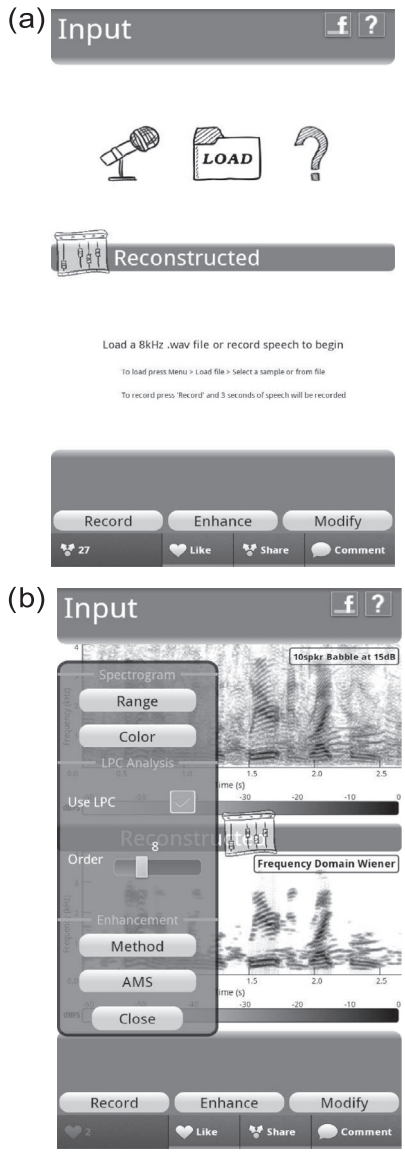


Fig. 2. UI activities presented to the user: (a) the main UI, demonstrating the look and feel once SEA is loaded, (b) the quick edit menu, allowing the user to change all configurable settings.

### 4.2. Other interactive content

In addition to recording, enhancement, modification and frame analysis at the touch of a button, SEA also incorporates interactivity by allowing each spectrogram to respond in the same manner as a button upon a user touch event.

#### 4.2.1. Adding noise

Upon the acquisition of an input speech signal and the generation of the spectrogram, the user can touch the spectrogram to access a range of options. Fig. 3(a) shows the options presented to the user. Once 'Add noise' is selected, the user is presented with Fig. 3(b) showing the noise types available. Once selected Fig. 3(c) allows the user to select the desired signal to noise ratio (SNR). Refer to Section 9 for details on SNR calculation within SEA. After the level

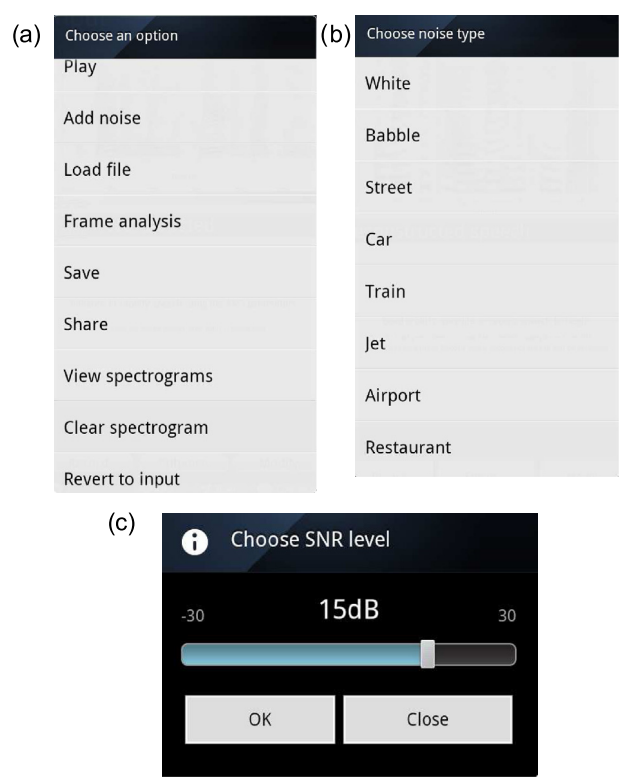


Fig. 3. Spectrogram touch event: (a) options first presented when the a touch event occurs on a spectrogram; (b) the additive noise sources available; (c) the SNR at which to add the noise.

of noise is selected the noisy speech is calculated and a new spectrogram is generated.

#### 4.2.2. Viewing the spectrograms

Once the spectrogram is generated, the user can also select 'View spectrogram' to open the spectrogram image in their desired image viewing application, providing the ability to zoom and if supported edit the generated image.

#### 4.2.3. Playing audio

The user can select to play audio attributed to any spectrogram that is generated.

#### 4.2.4. Sharing audio and the spectrogram

At any point, the user can share either the audio or spectrogram via supported social media, e-mail services or file transfer methods such as: Bluetooth, text messages, personal e-mail, Facebook, Google+ or Dropbox. This is useful in a class room environment to send students the work presented by the lecturer. In a laboratory environment where the student does not own a smart phone or tablet, they can send their own work accomplished by a laboratory device to themselves for their report. It also enables concepts to be communicated internationally at the push of a button; this feature could be useful for research students. All spectrogram and short-time graphs presented in this paper are automatically generated by SEA.


Fig. 4. Help menu designed to be interdisciplinary, encompassing help information for audio processing, speech processing and advanced filtering topics with links and references.

#### 4.2.5. User support

SEA provides four levels of technical support for users having difficulty in using the platform. Each level provides information on the use of SEA, theoretical principals and important resources. The three levels are

1. Comprehensive help menu available by pressing the question mark to the top right of Fig. 2(a). In this menu the user can choose help options depending on their objective, shown in Fig. 4.
2. A website which shows videos of performing certain function inside SEA, along other useful resources.[1]
3. This document, as it contains a relatively in-depth look into the teaching content.
4. Ability to contact the developer directly to ask questions.

The first three options allow the student work their problem out independently, referring them to known research and academic literature to seek further advice on theoretical problems.

## 5. Analysis–modification–synthesis

SEA uses a short-time Fourier analysis–modification–synthesis (AMS) framework to decompose a speech signal into its short-time magnitude and phase spectral components. The two primary elements of SEA which utilise this framework are for speech enhancement and speech modification. This section, firstly describes the AMS procedure used for traditional speech processing applications and secondly how it has been adapted to suit the SEA educational platform.

### 5.1. AMS method

The short-time Fourier AMS framework typically consists of three stages: (1) the analysis stage, where the input speech is processed using STFT analysis[2]; (2) the modification stage, where either the magnitude or phase spectrum undergoes some kind of modification; and (3) the synthesis stage, where the inverse STFT is followed by overlap-add synthesis to reconstruct the output signal.

The analysis stage applies STFT analysis to a discrete-time input signal to produce the complex frequency spectrum $X(n,k)$. Fig. 5 shows the traditional AMS framework. For a discrete-time signal $x(n)$, the STFT is given by

$$X(m,k) = \sum_{n=-\infty}^{\infty} x(n)w(n-m)\mathrm{e}^{-i2\pi kn/N}, \tag{1}$$

[1] Website: www.enhancementapp.com.

[2] Note: In SEA, the Fast Fourier Transform algorithm is used to compute the DFT of each analysis frame using Java's libGDX package.

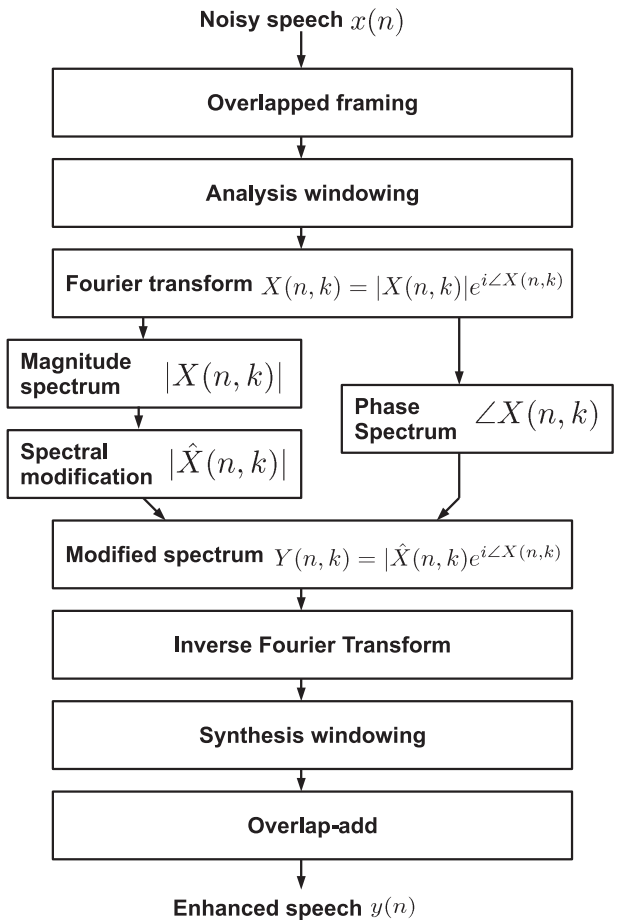


Fig. 5. Block diagram of a traditional AMS-based speech enhancement framework.

where $n$ refers to the discrete-time index, $k$ is the index of the discrete frequency, $N$ is the frame duration (in samples), and $w(n)$ is the analysis window function. In speech processing, a frame duration of 20–40 ms is typically used, with a Hamming window used as the analysis window function (Picone, 1993; Huang et al., 2001).

In polar form, the STFT of the speech signal can be written as

$$X(n,k) = |X(n,k)|\mathrm{e}^{i\angle X(n,k)}, \tag{2}$$

where $|X(n,k)|$ denotes the short-time magnitude spectrum and $\angle X(n,k)$ denotes the short-time phase spectrum.[3] Traditional AMS-based speech processing methods typically use a window of duration 32 ms (Oppenheim et al., 1979; Berouti et al., 1979; Martin, 1994; Virag, 1999). In the modification stage of the framework, the phase, magnitude or both spectra can be modified depending on the objectives. For speech enhancement purposes, it is common to enhance only the magnitude spectrum while keeping the phase spectrum unchanged. For example, given a noisy signal, $x(n)$, the modified complex frequency spectrum is given by a combination of the enhanced magnitude spectrum $|\widehat{X}(n,f)|$ and the noisy phase spectrum $\angle X(n,k)$, seen in polar form as

$$Y(n,k) = |\widehat{X}(n,k)|\mathrm{e}^{i\angle X(n,k)}. \tag{3}$$

Similarly, for modification purposes, the objective may be to perform some modification on the phase spectrum only, while keeping the magnitude spectrum unchanged. For example, given a clean input signal $s(n)$, the modified complex frequency spectrum is given by a combination of the clean magnitude spectrum $|S(n,k)|$ and the modified phase spectrum $\angle\widehat{S}(n,k)$, seen in polar form as

$$Y(n,k) = |S(n,k)|\mathrm{e}^{i\angle\widehat{S}(n,k)}. \tag{4}$$

Finally, the synthesis stage reconstructs the modified speech, $y(n)$, by applying the inverse STFT to the modified spectrum, followed by least-squares overlap-add synthesis (Quatieri, 2002).

### 5.2. Default configuration

On initial start up, the default AMS configuration is as follows: frame duration, $t_w$ of 32 ms with a 4 ms shift, and an DFT analysis length of $2N$ (where $N = t_w F_s$, and $F_s$ is the sampling frequency of input stimuli). The Hamming window is used as the analysis window.

### 5.3. Configurable settings

The user can configure all AMS parameter to examine relationships between frame duration, frame shift, analysis window, synthesis window and DFT length with speech quality and intelligibility. Frame duration ranges from 10 ms to 1023 ms, while the frame shift ranges from 2 ms to 250 ms. DFT analysis length can be changed from 128–8192 in powers of 2 depending on the desired effect. However, fail safes are in place which ensure SEA does not consume to much of the operating systems resources. These fail safes include:

- For large frame durations above 200 ms, DFT length is defined as the next power of two above the number of samples.
- For large frame durations above 200 ms, frame shift is fixed to 10% of the frame duration.
- If the selected DFT length is smaller than the signal length, DFT length is incremented to the next power of two above the signal length.

Rectangular window

$$w(n) = \begin{cases} 1, & 0 \leqslant n < N-1, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Hamming window

$$w(n) = \begin{cases} 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), & 0 \leqslant n < N-1, \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

[3] In the remainder of this paper, when referencing the magnitude and phase spectra the STFT modifier will be implied.

Hanning window

$$w(n) = \begin{cases} 0.5 - 0.5\cos\left(\frac{2\pi n}{N}\right), & 0 \leqslant n < N-1, \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

Blackmann window

$$w(n) = \begin{cases} 0.42 - 0.5\cos\left(\frac{2\pi(n-1)}{(N-1)}\right) + 0.08\cos\left(\frac{4\pi(n-1)}{(N-1)}\right), & 0 \leqslant n < N-1, \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

- Length of the signal is always zero padded to the DFT length.

The analysis and synthesis window functions can also be changed. The following window functions are available:

## 6. Loading speech and displaying the spectrogram

The first element SEA addresses is the loading of speech, calculation of the spectral data and displaying of the spectrogram.

### 6.1. Loading of speech

An input speech segment can be loaded in one of three ways: (1) Load from an assortment of pre-packaged, pre-recorded files of male and female speakers, (2) Load a file directly from the internal or external storage device, or (3) Record three seconds of speech directly from the devices microphone. The pre-packaged audio files are taken from the NOIZEUS speech corpus (Hu and Loizou, 2007).

### 6.2. Spectrogram

The spectrogram data was calculated using a fixed window size and frame shift of 32 ms and 4 ms respectively, and the Hamming window as the analysis window (see Eq. 6). After short-time Fourier analysis was conducted, the spectrogram was calculated as follows

$$specdata(n,k) = 20\log_{10}(|X(n,k)|). \quad (9)$$

Once the spectrogram data is computed, it is passed to a drawing routine to draw the spectrogram given the desired dimensions (device dependent). Fig. 6 illustrates two spectrograms drawn directly by SEA. The user can change the dynamic range of the spectrogram in the configuration menu, by setting a floor to the spectrogram at a desired dB below the maximum present in the signal, the default value is −60 dB.

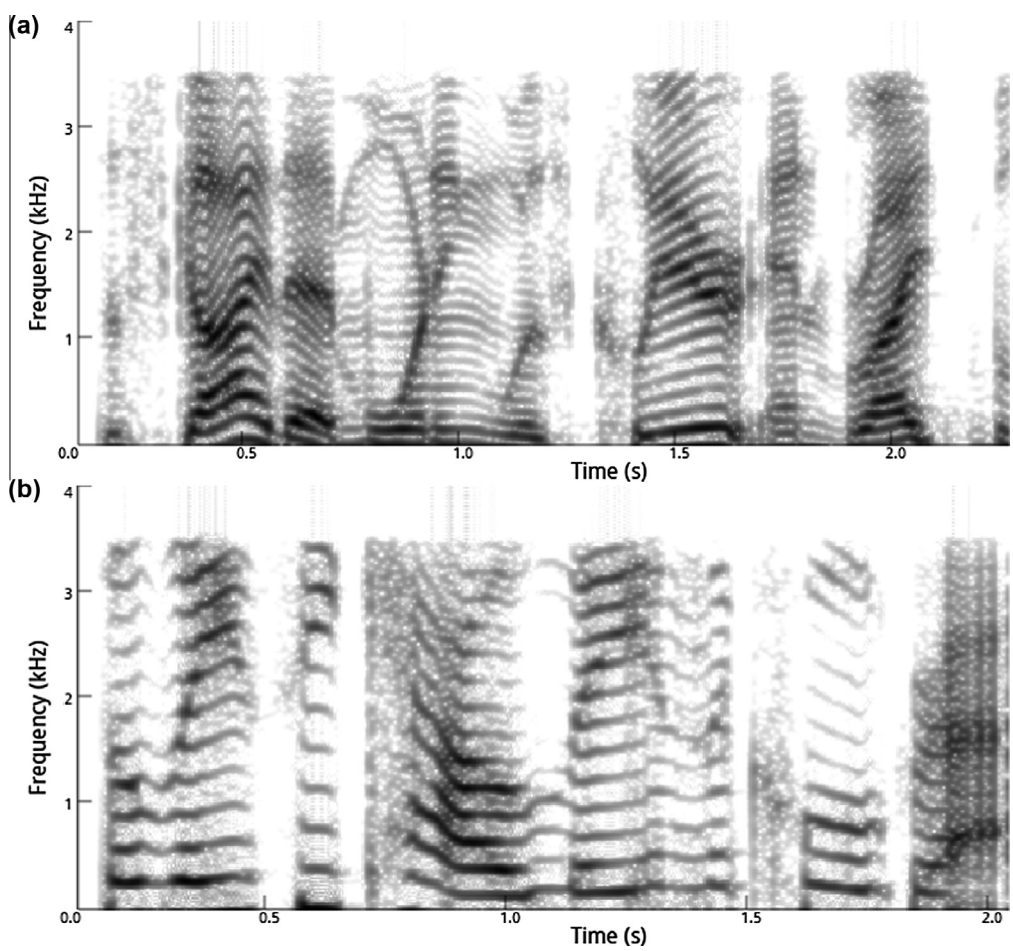


Fig. 6. Spectrogram created from speech, image generated by SEA; (a) male speaker, (b) female speaker.

## 7. Short-time analysis

Once a speech file is loaded, SEA provides a frame analysis feature, allowing the user to examine the spectrogram frame by frame. In this mode, the user can select a frame by moving a cursor across the generated spectrogram and viewing the corresponding data for that frame both in the time and frequency directions. Fig. 7(b) shows the UI for this mode, while Fig. 8 displays the images created by SEA, allowing the user to share their results with fellow students or teachers.

## 8. Linear Predictive Coefficients (LPCs)

LPCs can be a powerful tool for speech processing, namely in their ability to separate glottal pulse train information from articulatory filter influences, providing a spectral envelope where slow varying peaks throughout the spectrum can be closely associated to formant frequencies. SEA allows students to quickly calculate the LPCs for a loaded speech signal. This section will briefly describe LPCs and how to calculate them in SEA.

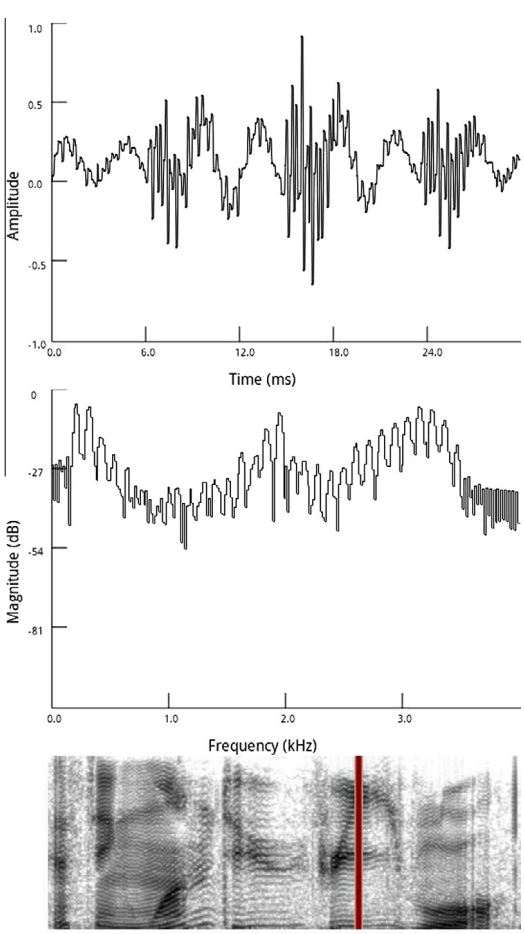


Fig. 7. UI for frame analysis tool; top: time domain frame, middle: magnitude spectrum frame, bottom: spectrogram cursor.

### 8.1. Calculating LPCs

Linear predictive (LP) analysis, is a fast, efficient method for estimating parameters of speech. All methods derived from Linear Predictive Coding revolves around a central process, which is that the current sample is estimated from a weighted linear combination of the samples that preceded it (Acero et al., 2001), illustrated in Eq. 10.

$$\hat{x}(n) = \sum_{k=1}^{p} a_k x(n-k) \tag{10}$$

where $\hat{x}(n)$ is the estimated value of the signal $x(n)$, and is equal to a combination of $p$ past samples where $a_k$ is the $k$th linear prediction coefficient which results in an inherently stable model. The optimum linear prediction coefficients are found in the minimum mean sense using the autocorrelation method. Where the error between the estimate of the signal and the actual signal is given in Eq. 11

$$e(n) = x(n) - \sum_{k=1}^{p} a_k x(n-k) \tag{11}$$

Minimising the prediction error by optimizing the prediction coefficients, $a_k$ is then done by solving a linear set of equations

$$\sum_{k=1}^{p} a_k R_{i-k} = R_i, \quad 1 \leqslant i \leqslant p \tag{12}$$

called the Yule–Walker equations. Where the autocorrelation function, $R_i$ is defined by

$$R_i = \sum_{n=0}^{N-1-i} x(n)x(n+i). \tag{13}$$

Levinson–Durbin method of solving this set of linear equations is very efficient and is commonly used (Makhoul, 1975)

Analysing the composition of this model can also be done in the frequency domain by taking the $z$-transform, so Eq. 11 becomes

$$E(z) = X(z)\left(1 - \sum_{k=1}^{p} a_k z^{-k}\right) = X(z)H(z)$$

where $X(z)$ is assumed to be convolved with the linear prediction model $H(z)$, so if $X(z)$ is the transfer function for speech, it can then be represented by an all pole model of speech (Milner, 2002), $\widehat{X}(z)$by transposing becomes

$$\widehat{X}(z) = \frac{G}{H(z)} = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}} \tag{14}$$

where $G$ is the model gain, given by

$$G^2 = R_0 - \sum_{k=1}^{p} a_k R_k$$

$$G = \sqrt{G^2}$$

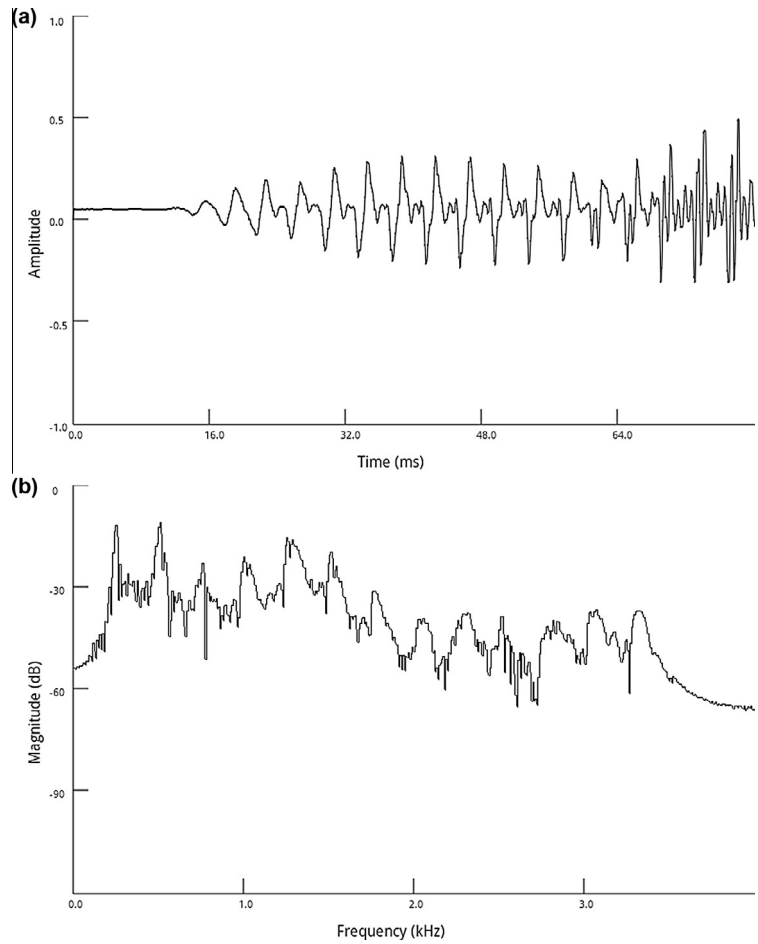


Fig. 8. Frame analysis images generated by SEA for 80 ms window duration, 8 ms frame shift and the hamming window used for analysis settings; (a) time domain representation, (b) magnitude domain representation in dB.

To analyze in the frequency domain, $z$ is then set to $e^{i\omega t}$ and taking the magnitude to give the resulting power spectrum.

$$\widehat{P}(\omega) = \frac{G^2}{|\,1 - \sum_{k=1}^{p} a_k \mathrm{e}^{-ik\omega T}|^2}. \tag{15}$$

SEA calculates the LPCs frame by fame, allowing the user to analyse the complete LPC spectrogram or just the LPC magnitude spectrum for a specific frame with the frame analysis tool. Fig. 9 shows an example of the LPC spectrogram generated by SEA and below that is the LPC magnitude spectrum for a voiced frame.

LPC analysis in SEA is accessed through the quick edit menu under "LPC Analysis", once enabled the student can select the order of the LP model, and choose which spectrogram to calculate the LPCs for (loaded or modified/enhanced).

## 9. Adding noise: SNR calculation

In Section 4.2.1 the process to add noise in SEA was discussed. This section further details the procedure which SEA uses to add noise at a desired SNR. When a user has selected the type of noise (white, babble, street, car, train, jet, airport or restaurant) they wish to add to their recording, they are asked to specify the SNR in dB. The noise signals come from the NOIZEUS speech corpus (Hu and Loizou, 2007), and are stored with SEA. Noise is added to the recorded speech using the following procedure:

1. Variance of the recorded speech is calculated as

$$\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x(n) - \mu_x)^2 \tag{16}$$

where $x$ is the recorded speech, $n$ is the sample number, $N$ is the recorded speech length and $\mu_x$ is the mean of the recorded speech.

2. Variance of the noise signal is calculated as

$$\sigma_d^2 = \frac{1}{N}\sum_{i=1}^{N}(d(n) - \mu_d)^2 \tag{17}$$

where $d$ is the noise signal, $n$ is the sample number, $N$ is the length of the noise and $\mu_d$ is the mean of the noise.

3. Standard deviation of the noise signal is calculated as

$$\sigma_d = \sqrt{\sigma_d^2}. \tag{18}$$

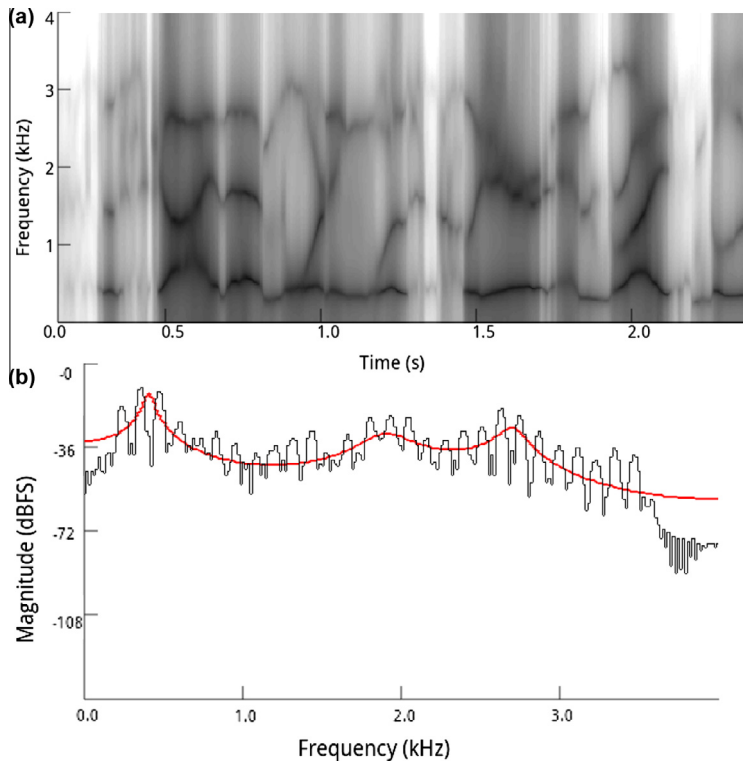


Fig. 9. LPC analysis in SEA; (a) LPC spectrogram using 8 coefficients, (b) LPC magnitude and acoustic magnitude spectra of a single voiced 32 ms frame within the spectrogram.

4. The noise signal is normalised to have a standard deviation of 1,

$$\tilde{d}(n) = \frac{d(n)}{\sigma_d}, \tag{19}$$

where $\tilde{d}$ is the normalised noise signal.
5. Signal to noise ratio, $r$ is calculated as

$$r = 10^{\mathrm{SNR}/10}, \tag{20}$$

where SNR is the desired SNR in dB, as chosen by the user.
6. Finally, the resulting noisy speech is obtained as

$$y(n) = x(n) + \left(\sqrt{\frac{\sigma_x^2}{r}}\right)\tilde{d}(n), \tag{21}$$

where $y$ is the resulting noisy signal.

However, one inherent problem with using this procedure alone is caused by Eq. 16 in step 1. In SEA, the recorded speech signal may have considerable durations of silent segments, adversely biasing SNR calculation. To account for this, SEA only calculates the noise level based on regions of speech present in the recording. Silence regions, or in this case silence frames (32 ms in duration) were classified as being silence if the power was $\leqslant$40 dB below the maximum power present in the recorded signal. Once silence frames were detected, they were omitted and the above 6 steps were followed to scale the additive noise signal according to speech-only frames.

Please note that the SNR definition used in this paper is ideally suited for speech corrupted by white noise. When noise is colored, this definition may not be that relevant. In such cases, it might be useful to introduce frequency weighting for the computation of SNR.[4] Since we encounter different kinds of colored noise in practice it is not possible to fix the frequency weighting for one particular kind of colored noise. However, if needed the user can add noise by using a frequency weighted SNR calculation, where they can decide about the shape of the weighting depending on the type of the colored noise which was added or recorded.

## 10. Role of magnitude and phase spectrum in speech

The role of the magnitude and phase spectrum for speech intelligibility has been the topic of many research papers (Ohm, 1843; Hermann von Helmholtz, 1954; Oppenheim et al., 1979; Wang and Lim, 1982; Paliwal and Alsteris, 2003; Alsteris and Paliwal, 2004; Alsteris and Paliwal, 2006; Shi et al., 2006; Wójcicki and Paliwal, 2007). In these studies it was shown that the magnitude spectrum is known to contribute significantly more to the intelligibility of speech then the phase spectrum when a 20–40 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

[4] For a more detailed discussion about the use of Frequency weighting when computing the SNR, please refer to: http://dnt.kr.hsnr.de/download/snr_comments.html.

when a low dynamic range synthesis window, i.e., rectangular window, with $\sim 1$ s duration is employed for speech analysis. 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. The two modification techniques provided are as follows:

1. Unit magnitude; and
2. Random phase.

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. This section will outline the procedure undertaken to perform both magnitude only and phase only reconstruction techniques.

### 10.1. Magnitude only reconstruction

Upon passing the loaded speech into the AMS framework with the user configured analysis and synthesis settings described in Section 5, Phase modification is performed as follows:

$$Y(n,k) = |X(n,k)|e^{i\angle\phi(n,k)}, \tag{22}$$

where $|X(n,k)|$ is the original signal magnitude, $\angle\phi(n,k)$ is the modified phase spectrum and $Y(n,k)$ is the new complex signal. In the case of random phase, $\angle\phi(n,k) = \xi$, where $\xi$ is a random number generated between $0 - 2\pi$.

Fig. 10 illustrates magnitude only reconstruction by randomising the phase spectrum. In (b) the window length is set to the default length of 32 ms and a frame shift of 4 ms using the hamming analysis window, resulting in speech which is perceived as pitchless, monotonous and relatively un-voiced. However, by changing the window length to large values ($\sim 1$ s), the magnitude spectrum no longer contains any useful information, seen in (c). The resulting speech is perceived as a continuous reverberate sound, illustrated in the spectrogram by the spectral and time smearing.

### 10.2. Phase only reconstruction

Upon passing the loaded speech into the AMS framework with the user configured settings, magnitude modification is performed as follows:

$$Y(n,k) = |\widehat{X}(n,k)|e^{i\angle X(n,k)}, \tag{23}$$

where $|\widehat{X}(n,k)|$ is the modified magnitude, $\angle X(n,k)$ is the original phase spectrum and $Y(n,k)$ is the new complex signal. In the case of unit magnitude, $|\widehat{X}(n,k)| = 1$ for all discrete frequency bins. Fig. 11 illustrates phase only reconstruction by using unit magnitude modification. Again, in (b) the window length is set to the default length of 32 ms and a frame shift of 4 ms using the hamming analysis window, resulting in speech which is seen as useless for speech processing, and perceived as strong additive noise with some fluctuation wrapped around voiced segments. However, by changing the window length to large values ($\sim 1$ s), the phase only reconstruction begins to convey useful information, seen in (c). The resulting speech is perceived the same as the original speech with some white noise at lower energy.

### 10.3. Discussion

This section discussed how SEA can demonstrate to some extend the role the phase and magnitude spectra have in terms of their contribution to the intelligibility of speech. The results presented here are consistent with the results presented in the literature (Oppenheim and Lim, 1981; Alsteris and Paliwal, 2004). In a class room environment, this can be useful to give the students a better understanding of the Fourier phase and magnitude spectra and how changing their short-time processing characteristics also effects the perceived speech.

## 11. Speech enhancement implementation

The primary objective in the field of speech enhancement is to remove noise from a noisy speech signal without causing or introducing distortions in the resulting speech. A number of speech enhancement algorithms have been proposed in the literature to remove noise from noisy speech. These algorithms fall into three categories: (1) spectral-subtractive algorithms, (2) statistical-model-based algorithms and (3) subspace algorithms. Only AMS based spectral-subtractive and statistical-model-based algorithms are used in SEA. Spectral subtractive algorithms obtain an estimate of the clean signal by subtracting an estimate of the noise signal from the noisy speech, illustrated in Section 11.3, while, statistical-model-based algorithms aim to estimate a set of discrete Fourier transform (DFT) coefficients that represent the clean speech when only measurements of the noisy speech are observed, illustrated in Section 11.4. These sections illustrate what has been implemented using the configurable AMS framework presented in Section 5.

### 11.1. Noise: the enemy

Noise is a term used in speech enhancement to identify any undesirable signals which corrupt the quality, and in extreme cases, the intelligibility of a speech signal. The performance of speech enhancement algorithms vary depending on the SNR and statistical properties of noise signal. So, it becomes imperative to the performance of an algorithm to have an accurate estimate of the noise present

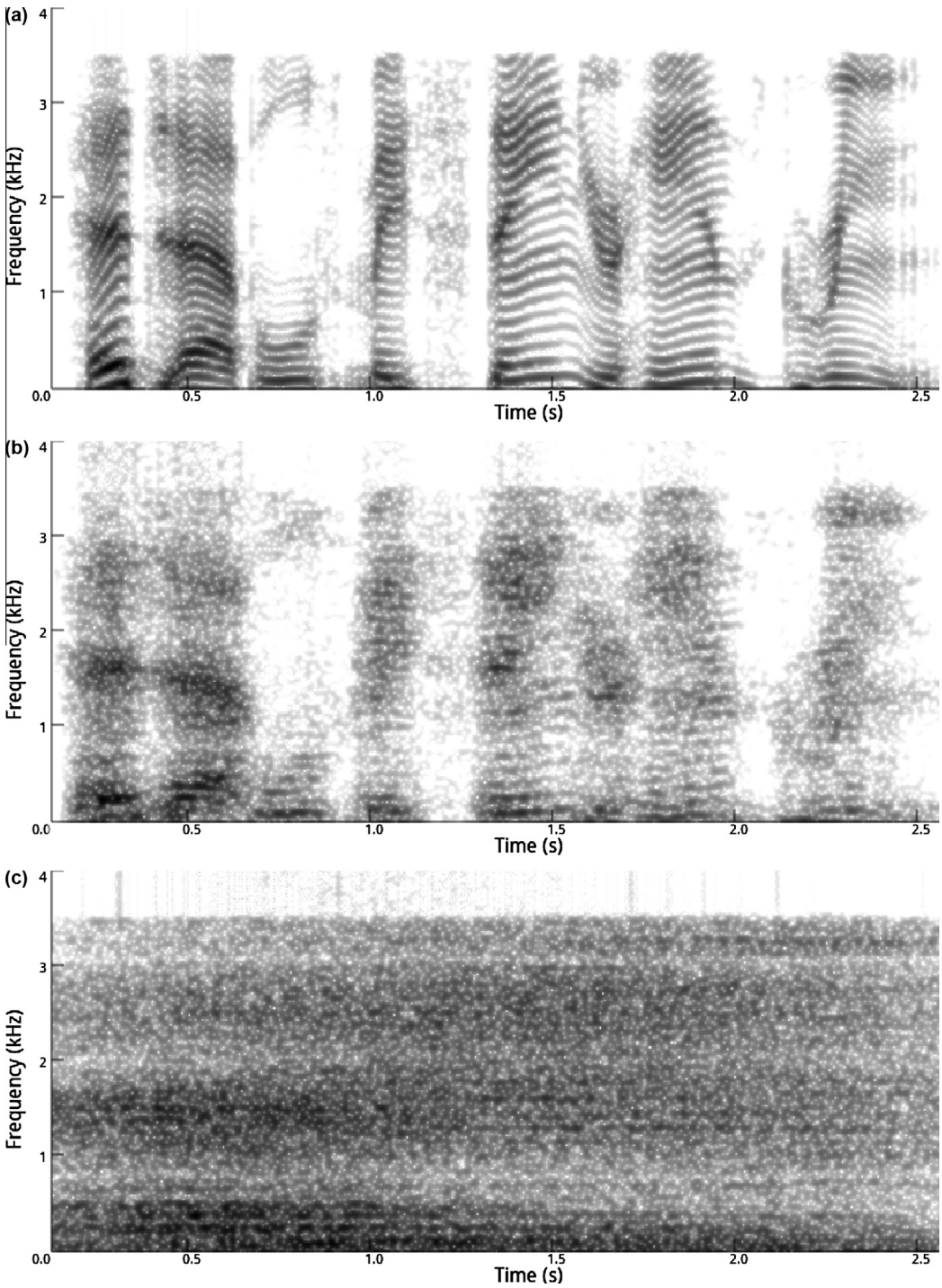


Fig. 10. Spectrograms for magnitude only reconstruction with random phase; (a) source speech, (b) default AMS settings, $tw$ =32 ms, $ts$ =4 ms and Hamming analysis window, (c) window duration increased to $\sim$ 1 s, $ts = 0.1tw$, Hamming analysis window.

in a noisy speech signal. If the noise estimate is too high, the speech may be affected, while if the it is too low, residual noise artifacts can become audible. A speech activity detector (SAD) is used in SEA to identify frames of speech and those of noise (Sohn et al., 1999). The SAD is a decision directed noise estimator which identifies speech present frames using a boolean operator. The SAD is based on the principal that noise is uncorrelated and adheres to the following additive noise model: Given an additive noise signal, $d(n)$ and a clean signal $x(n)$, the noisy signal can be considered a summation of the two as follows:

$$y(n) = x(n) + d(n). \tag{24}$$

By taking the STFT (see Eq. 1), Eq. 24 can be written in complex Fourier notation as

$$Y(i,k) = X(i,k) + D(i,k), \tag{25}$$

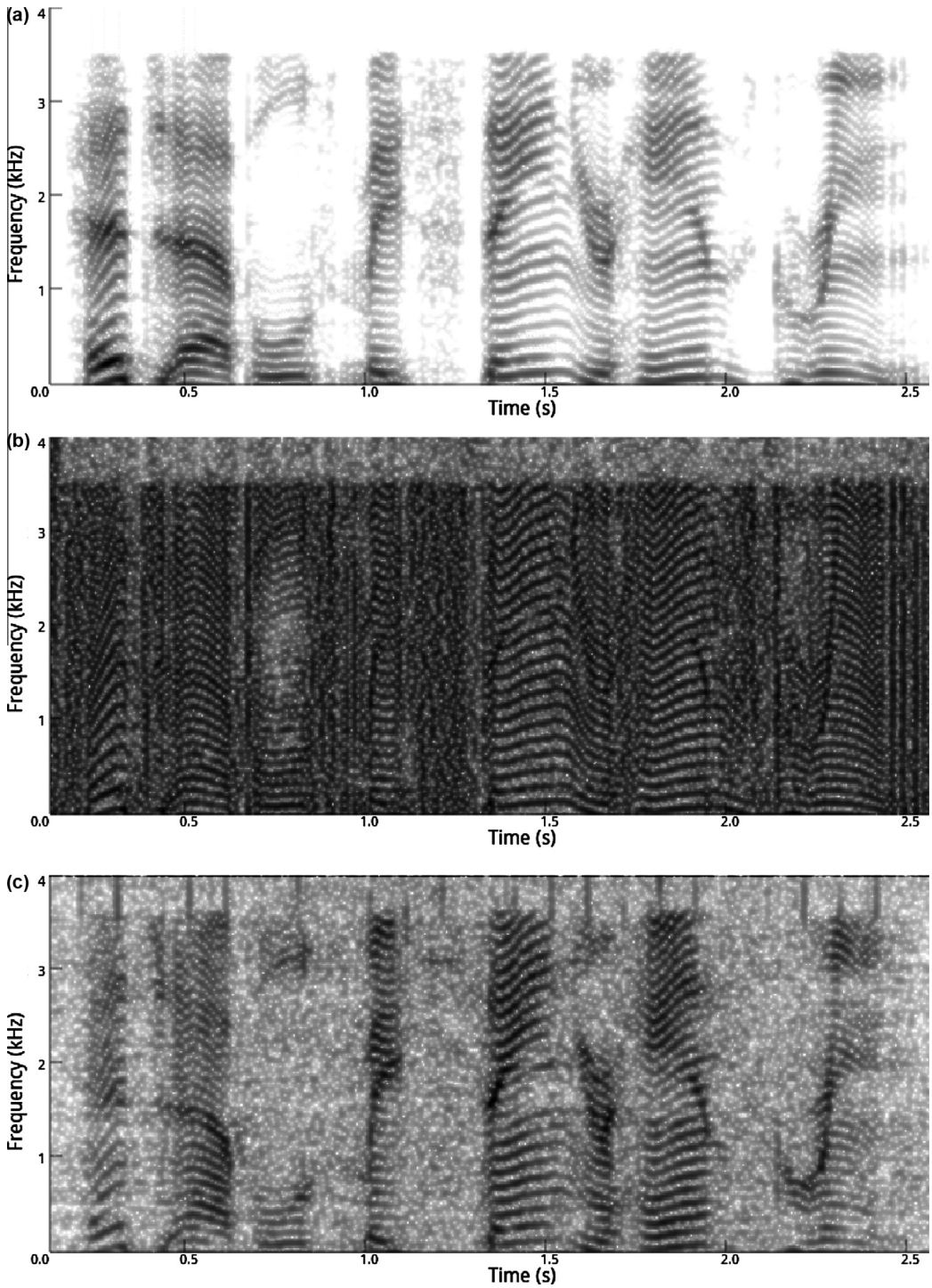


Fig. 11. Phase only reconstruction with unit magnitude; (a) source speech, (b) default AMS settings, $tw$ =32 ms, $ts$ =4 ms and Hamming analysis window, (c) window duration increased to $\sim$ 1 s, $ts = 0.1tw$, Hamming analysis window.

where $i$ refers to the frame index and $k$ is the discrete frequency index. The SAD defines two states or hypothesis for each frame within the noisy signal, $H_1$ as speech being present and $H_0$ as speech being absent

$$H_1 : \text{speech present} : Y(i,k) = X(i,k) + D(i,k), \tag{26}$$

$$H_0 : \text{speech absent} : Y(i,k) = D(i,k). \tag{27}$$

The SAD decision statistic is based off the geometric mean of the likelihood ratios (LRs)

$$\frac{1}{N}\sum_{k=1}^{N-1} \log\Lambda_k \underset{H0}{\overset{H1}{\gtrless}} \delta\delta \tag{28}$$

where $\delta = 0.15$ is a fixed speech presence threshold (Hu and Loizou, 2006), and

$$\Lambda_k = \frac{1}{1+\xi_k} \exp\left\{\frac{\gamma_k \xi_k}{1+\xi_k}\right\}, \tag{29}$$

here, $\gamma_k$ and $\xi_k$ are the a-posteri and a-priori SNRs given by

$$\gamma_k = \frac{|Y(i,k)|^2}{\lambda_d(k)}, \tag{30}$$

$$\xi_k = \alpha \frac{|\widehat{X}(i-1,k)|^2}{\lambda_d(i-1,k)} + (1-\alpha)P[\gamma_k - 1], \tag{31}$$

where $\lambda_d(k)$ is the expected power of the noise estimate for the $k$th spectral component, mixing constant $\alpha = 0.98$, and

$$P[x] = \left\{\begin{matrix} x, & \text{if } x \geqslant 0, \\ 0, & \text{otherwise.} \end{matrix}\right\} \tag{32}$$

In Eq. 31, $\xi_k$ is estimated recursively by the weighted sums of the previous amplitude estimate, $|\widehat{X}(i-1,k)|$ and the instantaneous SNR, $P[\gamma_k - 1]$ (Stark and Paliwal, 2011). When speech is absent ($H0 : Y(i,k) = D(i,k)$) the noise power estimate is updated by

$$|\widehat{D}(i,k)|^2 = (1-\beta)|Y(i,k)|^2 + \beta|\widehat{D}(i-1,k)|^2, \tag{33}$$

$$i = 2, 3, \ldots, I, \tag{34}$$

where $\beta = 0.98$, $|\widehat{D}(i-1,k)|^2$ is the previous noise estimate and $I$ is the number of frames (Loizou, 2007). For the first frame ($i = 1$) an initial estimate of noise is required. This estimate is taken from averaging the noisy signals power across the first 5 frames in pre-processing as it is assumed speech is not present using small frame durations. However, again this method is no longer desirable at larger window durations, and the noise estimate is no longer accurate resulting in poor noise removal. The noise estimate presented here is used for all speech enhancement algorithms implemented in SEA with the exception of spectral subtraction, where the noise estimate is taken only from the initial noise estimate.

### 11.2. SAD – limitations

The SAD was chosen for noise estimation because of its good performance in stationary noise environments, in addition SAD algorithms can be found in several known commercial applications, such as audio conferencing and GSM systems (Srinivasan and Gersho, 1993; Freeman et al., 1989). However, it is noted that a SAD may not be the ideal candidate for noise estimation in low-SNR environments, particularly when noise is non-stationary. Another consideration in the limitations of the SAD is frame duration (Loizou, 2007). The SAD performs well for speech enhancement in low SNR environments with a frame duration of 20–40 ms, however, due to the configurable nature of the AMS parameters, SEA allows the user to perform speech enhancement when large frame durations are selected. In this situation, any given frame may be non-stationary and can contain segments of speech absence and speech presence, often resulting in poor estimation of noise. Speech intelligibility is degraded significantly at large window durations, causing spectral smearing, reverberant noise and loss of speech. This aspect can help researchers gain an intuitive understanding on the relationship between the Fourier AMS parameters and the SAD, and their effects on perceived speech.

### 11.3. Spectral subtraction algorithm

In SEA, only one spectral subtraction algorithm has been implemented and is explained in this section. Here, SEA assumes the additive noise model presented in Eq. 25; however, the noise signal $D(i,k)$ is not a known quantity so this model becomes

$$Y(i,k) = X(i,k) + \widehat{D}(i,k) \tag{35}$$

where $\widehat{D}(i,k)$ is the estimated noise spectrum described in Section 11.1. Simple transposition of Eq. 35 can be done to extract a clean speech estimate

$$\widehat{X}(i,k) = Y(i,k) - \widehat{D}(i,k). \tag{36}$$

In practice, estimation is performed on the magnitude spectrum only, such that the complex Fourier representation of the enhanced speech becomes

$$\widehat{X}(i,k) = [|Y(i,k)| - |\widehat{D}(i,k)|]e^{i\angle Y(i,k)}, \tag{37}$$

where $\angle Y(i,k)$ is the noisy phase spectrum. Seen in Eq. 37 the resulting enhanced magnitude spectrum ($|\widehat{X}(i,k)| = |Y(i,k)| - |\widehat{D}(i,k)|$) can become a negative quantity if $|\widehat{D}(i,k)| > |Y(i,k)|$, to prevent this a basic half-wave-rectification process was implemented as follows

$$|\widehat{X}(i,k)| = \begin{cases} |Y(i,k)| - |\widehat{D}(i,k)|, & \text{if } |Y(i,k)| > |\widehat{D}(i,k)| \\ 0, & \text{otherwise.} \end{cases} \tag{38}$$

A well known shortcoming of this procedure is that inaccuracies in the noise estimate can result in erratic spectral peaks in each frame, which after the inverse Fourier transform and overlap add synthesis, result in tonal residual noise known as musical noise (Berouti et al., 1979). This can be easily demonstrated by adding stationary noise to speech at a low SNR and enhancing by spectral subtraction. The effects of spectral subtraction can be seen in Fig. 12, where white noise was added to a clean signal at 15 dB (shown in (b)) and enhanced with spectral subtraction using a frame duration, $t_w$ of 32 ms with a 4 ms shift and the hamming window as the analysis window. Fig. 12(c) shows the enhanced speech, where the remaining noise is musical in nature.

### 11.4. Statistical-model-based algorithms

Several other speech enhancement algorithms have been proposed in the literature based off measured statistical properties of the observed noisy signal. These methods have been specifically designed to improve the perceived quality of speech for a human listener and make some

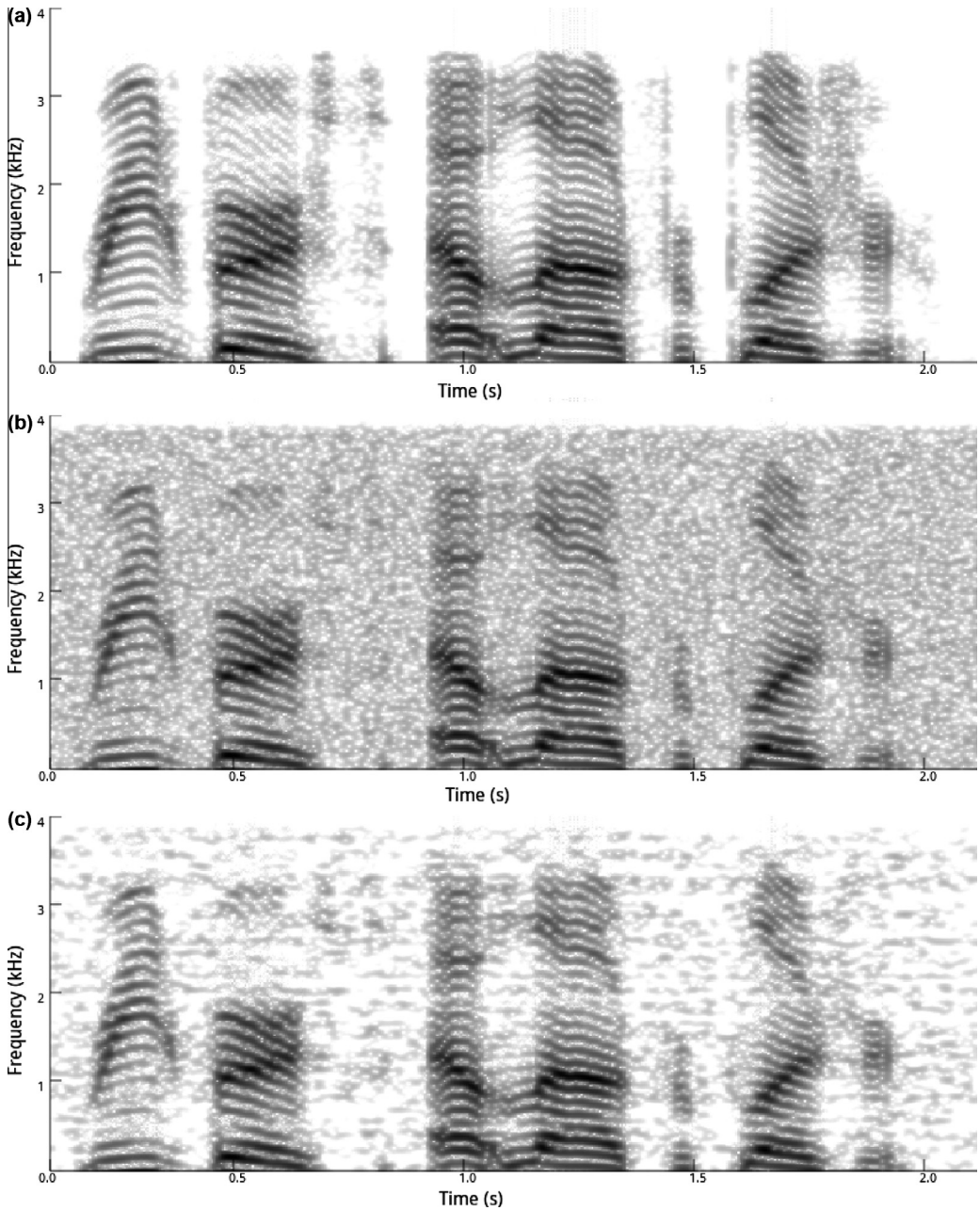


Fig. 12. Spectral subtraction method of speech enhancement; (a) clean speech, (b) speech corrupted with white noise at 15 dB and (c) enhanced speech.

underlying assumptions in order to derive an estimate of the clean speech signal. Firstly, the additive noise model presented in Eq. 25 is again assumed. Secondly, the STFT expansion coefficients $X(i,k)$ and $D(i,k)$ are assumed to be independent complex zero-mean Gaussian variables, with the expected power $\lambda_x(k) = E[|X(i,k)|^2]$ and $\lambda_d(k) = E[|D(i,k)|^2]$, where $E[\cdot]$ is the expectation operator. A detailed description of these assumptions can be found in (Ephraim and Malah, 1984). For typical AMS based speech enhancement, an estimate of the clean magnitude $(|\widehat{X}(i,k)|)$ is obtained from the noisy signal then reconstructed with the noisy phase as follows:

$$\widehat{X}(i,k) = |\widehat{X}(i,k)|e^{i\angle Y(i,k)} = Y(i,k) \cdot G(k), \tag{39}$$

where

$$G(k) = \frac{|\widehat{X}(i,k)|}{|Y(i,k)|} \tag{40}$$

is the spectral amplitude suppression function otherwise known as the gain function for the speech enhancement system. Gain functions are designed to attenuate specific frequencies of the noisy magnitude spectrum to estimate the enhanced magnitude spectrum. Several gain functions have been implemented, each with their own attenuation

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

| No. | Survey | sa(%) | a(%) | d | sd |
|---|---|---|---|---|---|
| 1. | This tool has helped me in understanding the purpose of speech enhancement and speech analysis | 63.64 | 36.36 | 0 | 0 |
| 2. | This tool easy to use | 68.18 | 31.82 | 0 | 0 |
| 3. | The meaning of the graphs and settings presented in this tool were logical and clear | 63.64 | 36.36 | 0 | 0 |
| 4. | This tool has contributed to my professional learning | 40.91 | 59.09 | 0 | 0 |
| 5. | This tool successfully links theory to practice | 54.55 | 45.45 | 0 | 0 |
| 6. | This tool was effective in helping me to develop an intuitive understanding | 59.09 | 40.91 | 0 | 0 |
| 7. | Presenting theoretical course material on my personal smart phone or tablet motivates me to learn | 72.73 | 27.27 | 0 | 0 |
| 8. | I would recommend similar teaching tools in other courses to help in the teaching process | 77.27 | 22.73 | 0 | 0 |
| 9. | This tool encourages me to further investigate speech processing | 50.00 | 40.91 | 9.09% | 0 |
| 10. | I enjoyed using this tool | 95.45 | 4.55 | 0 | 0 |
| 11. | I enjoy learning about speech processing and digital signal processing | 77.27 | 22.73 | 0 | 0 |
| | mean | 67.36 | 32.64 | 0 | 0 |

spectral shape. These gain functions are spectral Wiener (SW)(Loizou, 2007), Minimum Mean Square Error (MMSE) spectral amplitude (SA) (Ephraim and Malah, 1984), MMSE log-spectral amplitude (LSA) (Ephraim and Malah, 1985) and spectral energy (SE) and are provided below:

$$G_{SW}(k) = \frac{\xi_k}{1+\xi_k}, \tag{41}$$

$$G_{SA}(k) = \frac{\sqrt{\pi v_k}}{2\gamma_k}\exp\left(\frac{-v_k}{2}\right).\left[(1+v_k)I_0\left(\frac{v_k}{2}\right)+v_k I_1\left(\frac{v_k}{2}\right)\right], \tag{42}$$

$$G_{LSA}(k) = \frac{\xi_k}{1+\xi_k}\exp\left(\frac{1}{2}\int_{v_k}^{\infty}\frac{\exp(-t)}{t}dt\right), \tag{43}$$

$$G_{SE}(k) = \frac{\xi_k}{1+\xi_k}\sqrt{1+\frac{1}{v_k}}, \tag{44}$$

where

$$v_k = \frac{\xi_k}{1+\xi_k}\gamma_k. \tag{46}$$

The parameters $\xi_k$ and $\gamma_k$ are the *a priori* and *a posteriori* SNRs given by Eq. 31 and 30, respectively. $I_0(\cdot)$ and $I_1(\cdot)$ are given as the zeroth and first order modified bessel functions, respectively. In addition to these gain functions, the application of speech presence uncertainty (SPU) has been implemented for the MMSE SA estimator (Ephraim and Malah, 1984). SPU assumes a probabilistic two-state hypotheses; $H_1^k$ that speech is present for the $k$th spectral component or $H_0^k$ that speech is absent within the noisy spectrum. Incorporating this knowledge into the gain function improves the suppression of noise in a noisy signal particularly when the noise estimate is poor. The MMSE SA gain function with SPU is given by Loizou (2007):

$$G_{SA+SPU} = \frac{1-q_k}{1-q_k+q_k(1+\xi_k^*)\exp(-v_k^*)}\cdot G_{SA}, \tag{47}$$

where

$$\xi_k^* = \frac{\xi_k}{1-q_k}, \tag{49}$$

$$v_k^* = \frac{\xi_k^*}{1+\xi_k^*}\gamma_k \tag{50}$$

are the conditional SNRs assuming speech presence for the $k$th frequency bin, $q_k \triangleq p(H_0^k) = 0.3$ is the *a priori* probability of speech absence and $G_{SA}$ is the MMSE SA gain function given in Eq. 42. Appendix A demonstrates each speech enhancement algorithm operating in SEA by displaying the corresponding spectrograms.

### 11.5. Discussion

In this section, six known speech enhancement algorithms used in SEA have been outlined. SEA allows the user to modify the AMS parameters prior to enhancing a speech signal, allowing the investigation into the strengths and weaknesses of each method.

## 12. Student evaluation

### 12.1. Subjects and conditions

An online survey was conducted on twenty-two undergraduate students in a speech processing course at the University of Texas, Dallas after using SEA for a one-hour laboratory. The laboratory was included as a non-assessable compulsory item in the course and was conducted for only two students at a time in a quiet room over a period of three weeks toward the end of the teaching semester. The tasks on the laboratory were tailored to align with content taught within the course and covered topics such as speech production, speech as a stochastic signal, the role of the Fourier magnitude and phase spectra for speech intelligibility and speech enhancement.[5]

### 12.2. Procedure

The students were provided with a Samsung Galaxy tablet running the Android 3.2 operating system with SEA pre-installed. After completing the laboratory the students

[5] The laboratory script used in this work can be accessed for free at the following website: www.enhancementapp.com/resources/lab.

Table 2
Kruskal–Wallis ANOVA tables of survey questions, with post-hoc MannWhitney *U* analysis, where sa = strongly agree, a = agree, d = disagree and sd = strongly disagree.

| No. | Source | SS | df | MS | K | p | post-hoc |
|---|---|---|---|---|---|---|---|
| 1. | Between | 13861 | 3 | 4620.36 | 37.04 | < 0.05 | sa > d,sd |
| | Within | 18699 | 84 | 222.61 | | | |
| | Total | 32560 | 87 | | | | |
| 2. | Between | 15395 | 3 | 5131.5 | 41.18 | < 0.05 | sa > d,sd |
| | Within | 17127 | 84 | 203.89 | | | |
| | Total | 32522 | 87 | | | | |
| 3. | Between | 13861 | 3 | 4620.4 | 37.04 | < 0.05 | sa > d,sd |
| | Within | 18699 | 84 | 222.61 | | | |
| | Total | 32560 | 87 | | | | |
| 4. | Between | 10727 | 3 | 3575.7 | 28.64 | < 0.05 | a > d,sd |
| | Within | 21860 | 84 | 260.24 | | | |
| | Total | 32588 | 87 | | | | |
| 5. | Between | 11631 | 3 | 3877.1 | 31.04 | < 0.05 | sa > d,sd |
| | Within | 20973 | 84 | 249.68 | | | |
| | Total | 32604 | 87 | | | | |
| 6. | Between | 12599 | 3 | 4199.7 | 33.64 | < 0.05 | sa > d,sd |
| | Within | 19988 | 84 | 237.96 | | | |
| | Total | 32588 | 87 | | | | |
| 7. | Between | 17177 | 3 | 5725.8 | 46.02 | < 0.05 | sa > a,d,sd |
| | Within | 15295 | 84 | 182.08 | | | |
| | Total | 32472 | 87 | | | | |
| 8. | Between | 19188 | 3 | 6396.1 | 51.51 | < 0.05 | sa > a,d,sd |
| | Within | 13223 | 84 | 157.42 | | | |
| | Total | 32412 | 87 | | | | |
| 9. | Between | 87831 | 3 | 2927.7 | 23.40 | < 0.05 | sa > sd |
| | Within | 23876 | 84 | 284.24 | | | |
| | Total | 32659 | 87 | | | | |
| 10. | Between | 29098 | 3 | 9699.3 | 78.96 | < 0.05 | sa > a,d,sd |
| | Within | 29615 | 84 | 35.26 | | | |
| | Total | 32060 | 87 | | | | |
| 11. | Between | 19188 | 3 | 6396.1 | 51.51 | < 0.05 | sa > a,d,sd |
| | Within | 13223 | 84 | 157.42 | | | |
| | Total | 32412 | 87 | | | | |
| | | | | | | | *p< 0.05 |

were instructed to complete an online survey assessing their personal experience with SEA. The survey consisted of eleven questions based on a relatively equidistant 4-point Likert-type scale with the following conditions; strongly agree (sa), agree (a), disagree (d) and strongly disagree (sd); followed by an open ended question for students to provide further feedback. The decision to use a 4-point scale rather than a 5-point scale was made to simplify the process, although this type forces the student to make a preference instead of being undecided (Allen and Seaman, 2007); it has been shown the difference is negligible (Armstrong, 1987).

### 12.3. Results and discussion

Table 1 shows the average results of the student responses to the survey. Making the assumption that the data collected is non-parametric, the Kruskal–Wallis one-way analysis of variance refereed to here as (kANOVA) was run to asses the significant effect of each question and further non-parametric post-hoc MannWhitney *U* testing was also run to examine the significant effects between responses for each question (sa,a,d,sd). Table 2 shows the kANOVA and post-hoc analysis results. Each survey question indicated a significant effect for kANOVA ($p < 0.05$); post-hoc testing provided the specific significant differences between responses. In questions 1–3,5 and 6, "sa" was preferred significantly higher then "d" and "sd"; questions 7,8,10 and 11, "sa" was preferred significantly higher than all other responses. Question 4, students preferred "a" significantly higher then "d" and "sa"; question 9 students preferred "sa" significantly higher than "sd".

All questions indicated a significantly positive effect toward "sa" or "a", supporting the hypothesis that SEA is a useful, enjoyable platform which equips students with a quick and easy way to apply theory to practice. SEA was developed independently from the specific course con-

tent presented in the class these students were undertaking, none-the-less SEA's flexibility of course content was comprehensible and logical to the students.

It is noted that this was a short-term study on students personal experience with SEA over a one hour laboratory, where assessable learning outcomes were not obtainable. However, some pedagogical long-term learning likelihood's can be rationalised from the above results using cognitive evaluation theory (CET). CET describes the potential effects of external events on internal motivation (Deci and Ryan, 1985). Given that SEA (external event) was found to be enjoyable, encouraging, motivating and effective in aiding the students in gaining an intuitive understanding to a significant degree, it could be said SEA had a positive effect on perceived self-determination and competence and hence will enhance intrinsic motivation and learning (Deci et al., 2001). Although many factors (known and unknown) contribute to the acceleration of a students learning, the increase in the students perceived self-determination and competence is the first step in achieving this goal.

## 13. Teacher evaluation

The student evaluation provided an insight into the usefulness of SEA in a classroom and laboratory environment. And although it was not a long term study it provided some indicators regarding the potential to improve student engagement over time using CET. One element which the short-term test does not assess is the students attention level after continual use of SEA over a period of weeks. For this purpose, another survey was conducted on educational professionals who teach speech related DSP courses at universities across the world. Although, this to is not a long term study, teaching professional have experience in teaching with various tools over time, and their survey answers can provide a clearer picture on the usefulness of SEA as an additional learning tool in their own classes over time.

### *13.1. Subjects, conditions and procedure*

An online survey with eight questions was conducted on ten teachers of DSP related subjects from five different Universities around the world. The teachers were each asked to trial SEA and assess whether it would be a good teaching tool for students in their respective courses. Each teacher had access to an Android device and the survey could be conducted either on the device or on a PC with Internet access. The experiment took one-month to complete and the results were subjected to the same statistical analysis as the first experiment.

### *13.2. Results and discussion*

Table 3 shows the results of the teacher survey, where the results were collected using the same Likert-type scale as the student survey (sa,a,d,sd). Table 4 shows kANOVA table for each question along with the results of the post-hoc Mann-Whitney U test for each question. All questions indicated that a significant difference occurred between responses. And the post-hoc analysis results showed; questions 1–2,4 and 7, "a" was preferred significantly higher then "d" and "sd"; question 3, "sa" was preferred significantly higher then "sd"; question 8, "sa" was preferred significantly higher then "d" and "sd"; question 6, "a" was preferred significantly higher then "sd"; finally, the control question 5, "d" was preferred significantly higher then "a" and "sa".

With the exception of the control question (question 5), all other supporting questions showed a significant trend toward "sa" or "a" and question 5 the trend was toward "d". The result of question 5 indicated that student attention will not diminish over time while using SEA, at least no more than using any other teaching material like a textbook or laboratory notes. The results of the remaining questions indicated that the teachers believed students will even benefit from such an application, aligning with the results of from the student survey. Helping to increase engagement, enjoyment and motivation. Additionally, the results of question 8, indicated that the teachers agreed that an application on touch screen devices which are rapidly integrating into our daily lives are good interfaces to present and learn content from.

## 14. Course integration

SEA has been designed to be a complementary learning tool, assisting students to understand content taught in lectures through practical demonstrations and visualisations. It is intended to accompany students in class and during laboratories. If the student does not have a suitable Android device, the University could provide a tablet to students in pairs when conducting an exercise. SEA covers a broad range of topics covering DSP basics to advanced DSP for speech, and could be well suited for the following DSP theory structure:

1. Non-parametric spectral estimation: Short-time Fourier transform (Calculating the periodogram of a time sequence, frame by frame) and windowing, refer to Section 5.
2. Parametric spectral estimation methods, namely AR modeling with LPCs, refer to Section 7.
3. Spectrogram analysis of speech, encompassing the role of the Magnitude and Phase spectrum for speech quality and intelligibility, refer to Section 6 and Section 7.
4. Spectral subtractive techniques of noise removal, refer to Section 11.3.
5. Noise estimation: statistical model update with a SAD, refer to Section 11.1.
6. Optimum linear filtering in the minimum mean-square error (MMSE) sense, namely Wiener filtering. This topic can also cover FIR and IIR filtering, refer to Section 11.4.
7. Advanced statistical model based filtering methods, also refer to Section 11.4.

Table 3
Results of a teacher survey evaluating the usefulness of SEA, where sa = strongly agree, a = agree, d = disagree and sd = strongly disagree.

| No. | Survey | sa | a | d | sd |
|---|---|---|---|---|---|
| 1. | SEA is a valuable learning tool for speech enhancement and speech analysis | 30% | 70% | 0 | 0 |
| 2. | SEA is a useful tool which connects theory to practice for students | 40% | 60% | 0 | 0 |
| 3. | Presenting theory on a mobile device is appealing to students | 60% | 40% | 0 | 0 |
| 4. | SEA would be useful in your classroom to demonstrate theory | 30% | 70% | 0 | 0 |
| 5. | Students will get bored using SEA in conjunction with other teaching material | 0 | 0 | 60% | 40% |
| 6. | SEA will positively effect student engagement | 40% | 50% | 10% | 0 |
| 7. | Students will enjoy using SEA | 40% | 60% | 0 | 0 |
| 8. | Would you recommend more educational apps like SEA | 60% | 40% | 0 | 0 |

Table 4
Kruskal–Wallis ANOVA tables of teacher survey questions, with post-hoc MannWhitney $U$ analysis, where sa = strongly agree, a = agree, d = disagree and sd = strongly disagree.

| No. | Source | SS | df | MS | K | p | post-hoc |
|---|---|---|---|---|---|---|---|
| 1. | Between | 1510.05 | 3 | 503.35 | 19.29 | < 0.05 | a > d,sd |
| | Within | 1542.45 | 36 | 42.85 | | | |
| | Total | 3052.5 | 39 | | | | |
| 2. | Between | 1204.8 | 3 | 401.6 | 15.36 | < 0.05 | a > d,sd |
| | Within | 1855.2 | 36 | 51.53 | | | |
| | Total | 3060 | 39 | | | | |
| 3. | Between | 1012.8 | 3 | 337.6 | 12.91 | < 0.05 | sa > sd |
| | Within | 2047.2 | 36 | 56.87 | | | |
| | Total | 3060 | 39 | | | | |
| 4. | Between | 1510.05 | 3 | 503.35 | 19.29 | < 0.05 | a > d,sd |
| | Within | 1542.45 | 36 | 42.85 | | | |
| | Total | 3052.5 | 39 | | | | |
| 5. | Between | 1012.8 | 3 | 337.6 | 12.91 | < 0.05 | d > a,sa |
| | Within | 2047.2 | 36 | 56.87 | | | |
| | Total | 3060 | 39 | | | | |
| 6. | Between | 678.05 | 3 | 226.02 | 8.62 | < 0.05 | a > sd |
| | Within | 2389.45 | 36 | 66.37 | | | |
| | Total | 3067.5 | 39 | | | | |
| 7. | Between | 1204.8 | 3 | 401.6 | 15.36 | < 0.05 | a > d,sd |
| | Within | 1855.2 | 36 | 51.53 | | | |
| | Total | 3060 | 39 | | | | |
| 8. | Between | 1012.8 | 3 | 337.6 | 12.91 | < 0.05 | sa > d,sd |
| | Within | 2047.2 | 36 | 56.87 | | | |
| | Total | 3060 | 39 | | | | |
| | | | | | | | *p< 0.05 |

Each of the above teaching topics can be demonstrated on a real speech signal with SEA, both at the lecture and at the beginning of each laboratory session. SEA enhances the traditional "lecture-then-test" teaching scheme by bringing the "test" stage to the student immediately after or even during content presentation in class. Additionally, Lecturers with access to an Android tablet or phone with a mirroring HDMI output function can directly present the connection of theory to practice via a projection system, or could simply pass the device among the students. Another use for SEA would be to conduct a speech processing summary laboratory like the one used in Section 12, where toward the end of the teaching period students could conduct a one hour overview of speech processing theory prior to examination.

## 15. Limitations

Although, seen in this work, mobile computing power has increased and drastically transformed the way students can interact with course content, SEA does have limitations. These limitations are due to hardware and operating system restrictions and are outlined in Table 5. As Android OS is supported by many different manufactures who often create different models of devices with varying screen sizes, SEA application footprint in regards to memory usage changes per device due to its rendering capabilities on smaller and larger screen sizes. Additionally, in-line with many different manufactures, audio capturing hardware varies from device model to model, and hence compatibility with all devices becomes cumbersome.

Table 5
Hardware Limitation imposed on SEA on current Android devices.

| No. | Cause | Limitation |
|---|---|---|
| 1. | Processor | Speed |
| 2. | Memory | Heap space restrictions on the Android OS |
| 3. | Recording | Many devices with different hardware |

## 16. Future work and distribution

SEA is currently distributed on the Google Play market, available to all users with phones and tablets that use Android operating system 2.3.3 and above. There are two versions, one called SEA lite and SEA. SEA lite is free, and allows students access the majority of features with a few limitations on saving images and LPC analysis. However, SEA is the fully featured Android application with no restrictions and costs less then a coffee. The SEA project will be constantly improving, some of the planned improvements include:

- The addition of other spectral modification techniques.
- The addition of other speech enhancement methods.
- The addition of other analysis and synthesis window functions.
- Constant UI interface modifications to ensure a quality user experience.
- Optimisation to ensure speed and usability.

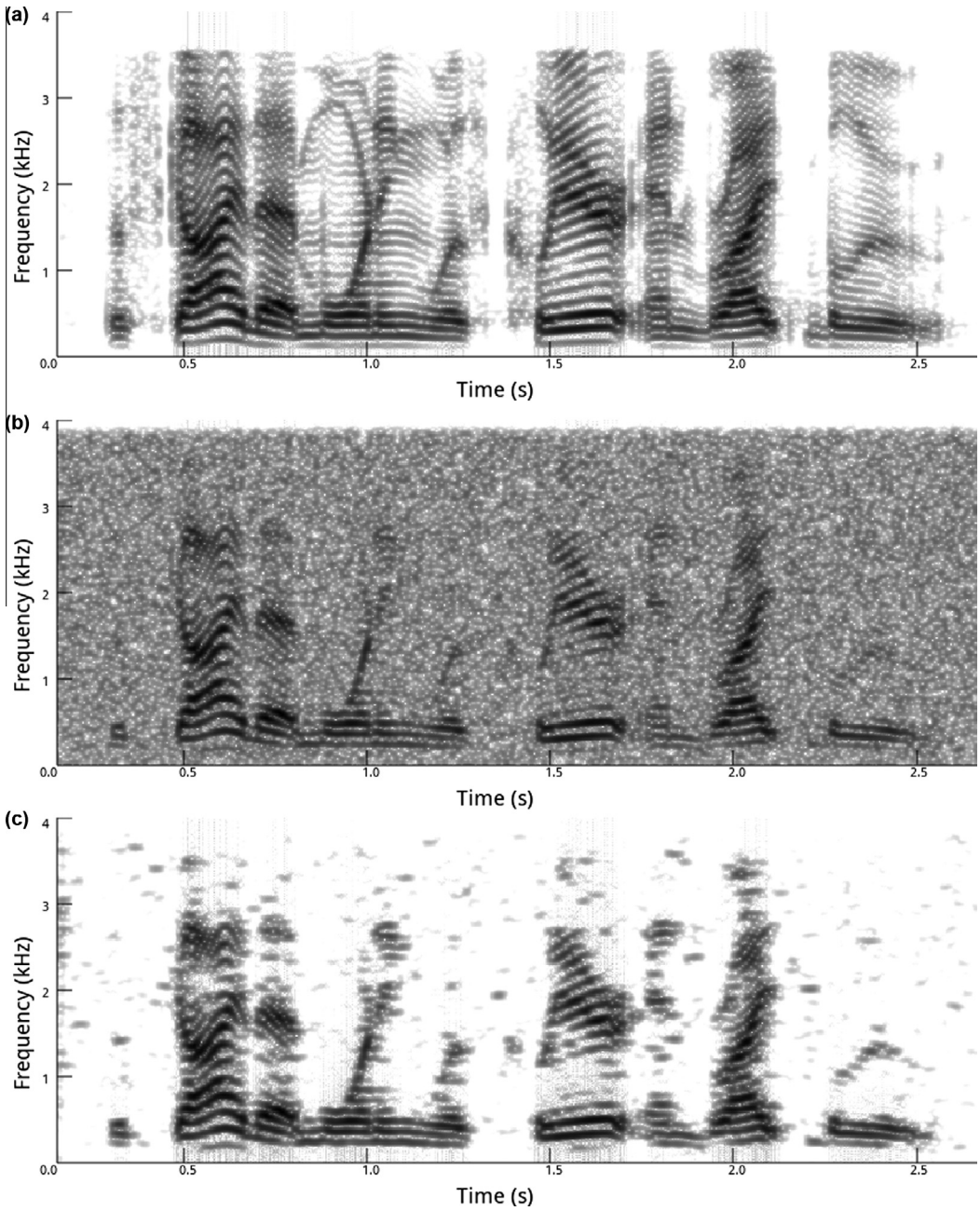


Fig. 13. Spectral Wiener method of speech enhancement; (a) clean speech, (b) speech corrupted with white noise at 15 dB and (c) enhanced speech.

The SEA project is open to suggestions from other professionals in the field to continually improve as hardware evolves. Deploying to other platforms such iOS and a stand alone PC environment is also being investigated.

## 17. Conclusion

Smart mobile devices are becoming readily available to university students, capitalising on their mobility, computing power and interactive capabilities can increase the effectiveness of teaching. This paper has suggested the use of a novel mobile educational tool to assist teaching DSP based speech processing content along side lectures and laboratories within a higher education curriculum. The tool is called Speech Enhancement for Android (SEA) and is available on the Google Play market for the cost of basic stationary. SEA was found to be significantly enjoyable and effective in motivating students in pursuing a higher level of competence in the comprehension of course content. Additionally, this work also provided an in-depth look at the content behind SEA acting as a road-map for lecturers, along with suggestions on how SEA can be incorporated into a current teaching framework.

## Acknowledgments

The authors would like to thank Professor P. Loizou and Dr K. Wójcicki for their support and time in helping

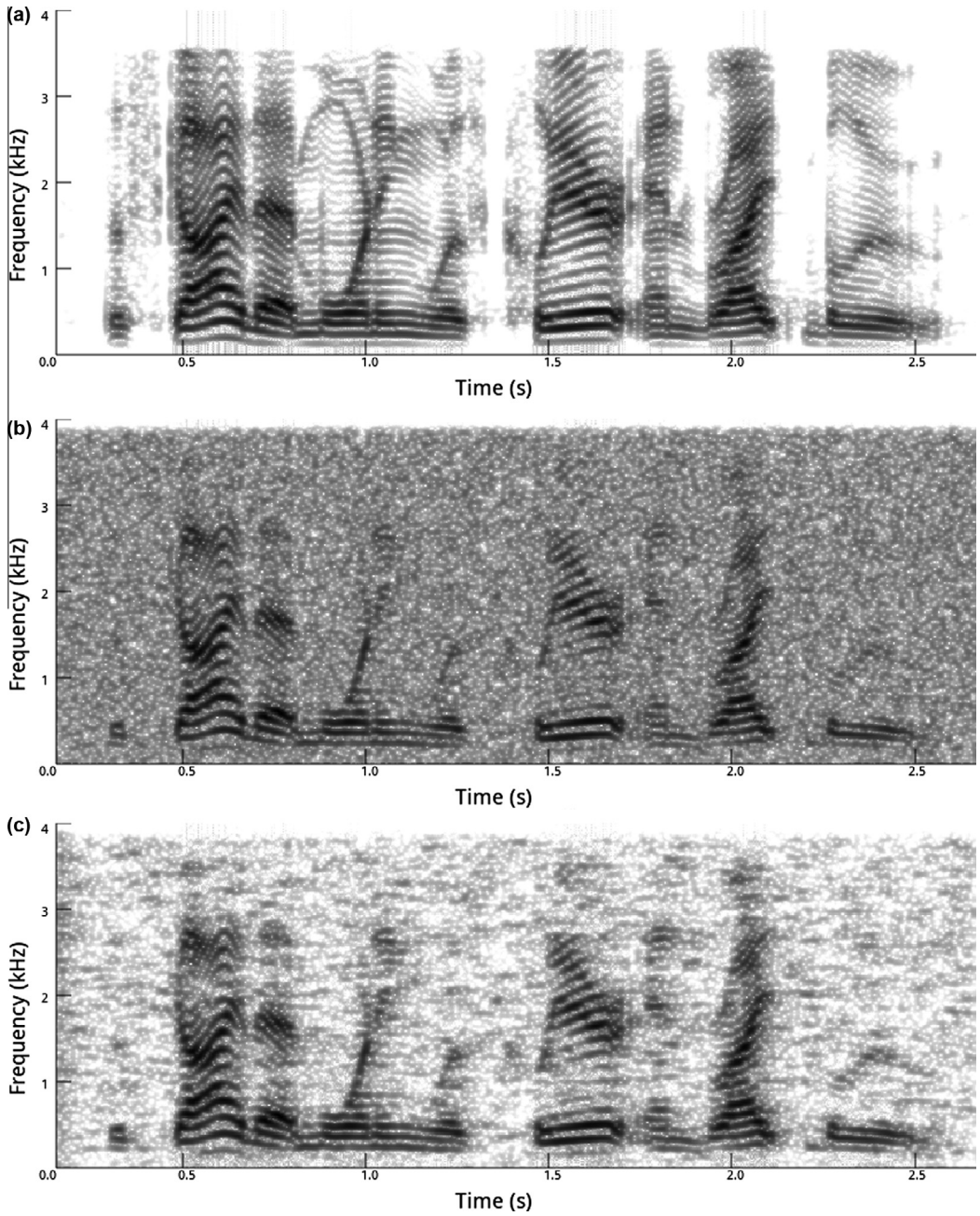


Fig. 14. MMSE SA method of speech enhancement; (a) clean speech, (b) speech corrupted with white noise at 15 dB and (c) enhanced speech.

to conduct the student survey at the University of Texas, Dallas. In memoriam of Professor P. Loizou and his contribution to speech processing research.

## Appendix A. Demonstration of speech enhancement algorithms through SEA

A few popular speech enhancement algorithms are used to generate the spectrograms shown in this appendix. The following settings were used. Additive white noise at 15 dB was used to corrupt the clean speech.

*AMS settings:* Window length = 32 ms, frame shift = 4 ms, analysis window = Hamming.

*Source audio:* sp10 from NOIZEUS speech corpus, male speaker "The sky that morning was clear and bright blue".

*SAD settings:* threshold =0.15, noise update rate = 0.98, smoothing factor = 0.98, a priori minimum = −25 dB.

*SPU settings (For MMSE SA +SPU):* a priori probably of speech absence = 0.30.

Fig. 13 shows the spectral wiener filter. Fig. 14 shows the MMSE spectral amplitude estimator. Fig. 15 shows the MMSE SA estimator with speech presence uncertainty (SPU). Fig. 16 shows the Log MMSE SA estimator. Finally, Fig. 17 shows the spectral energy estimator.

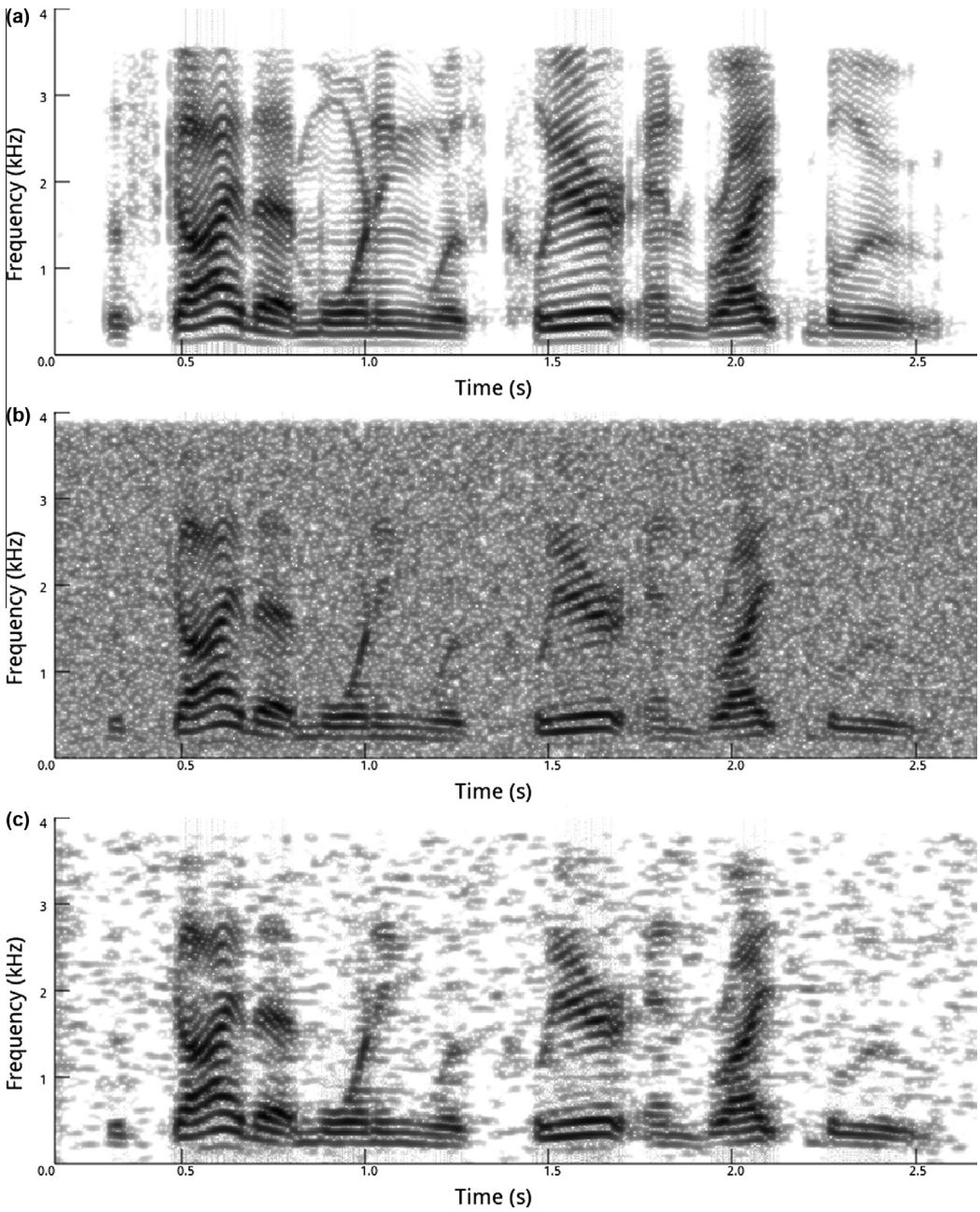


Fig. 15. MMSE SA + SPU method of speech enhancement; (a) clean speech, (b) speech corrupted with white noise at 15 dB and (c) enhanced speech.

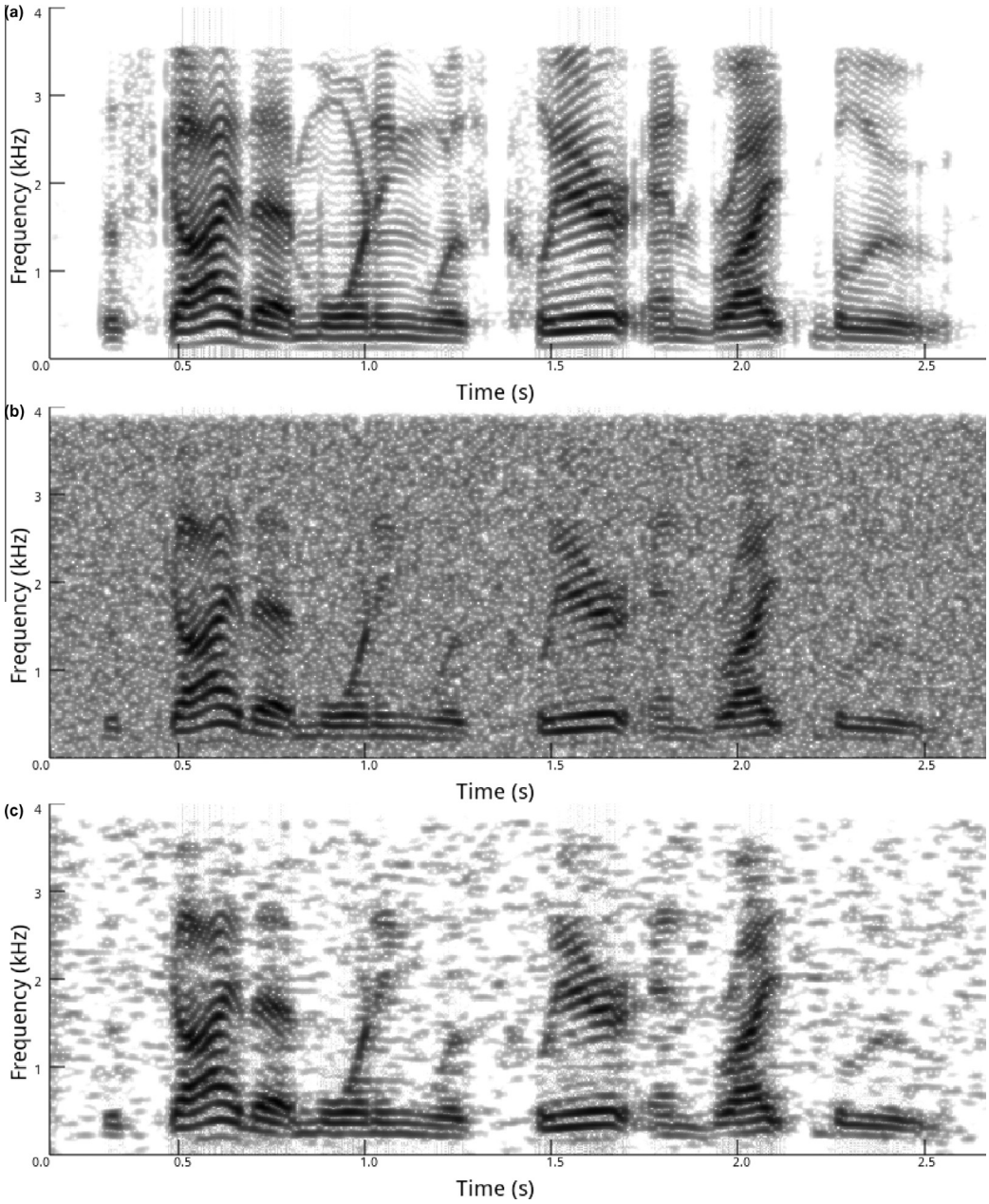


Fig. 16. Log MMSE method of speech enhancement; AMS settings: (a) clean speech, (b) speech corrupted with white noise at 15 dB and (c) enhanced speech.

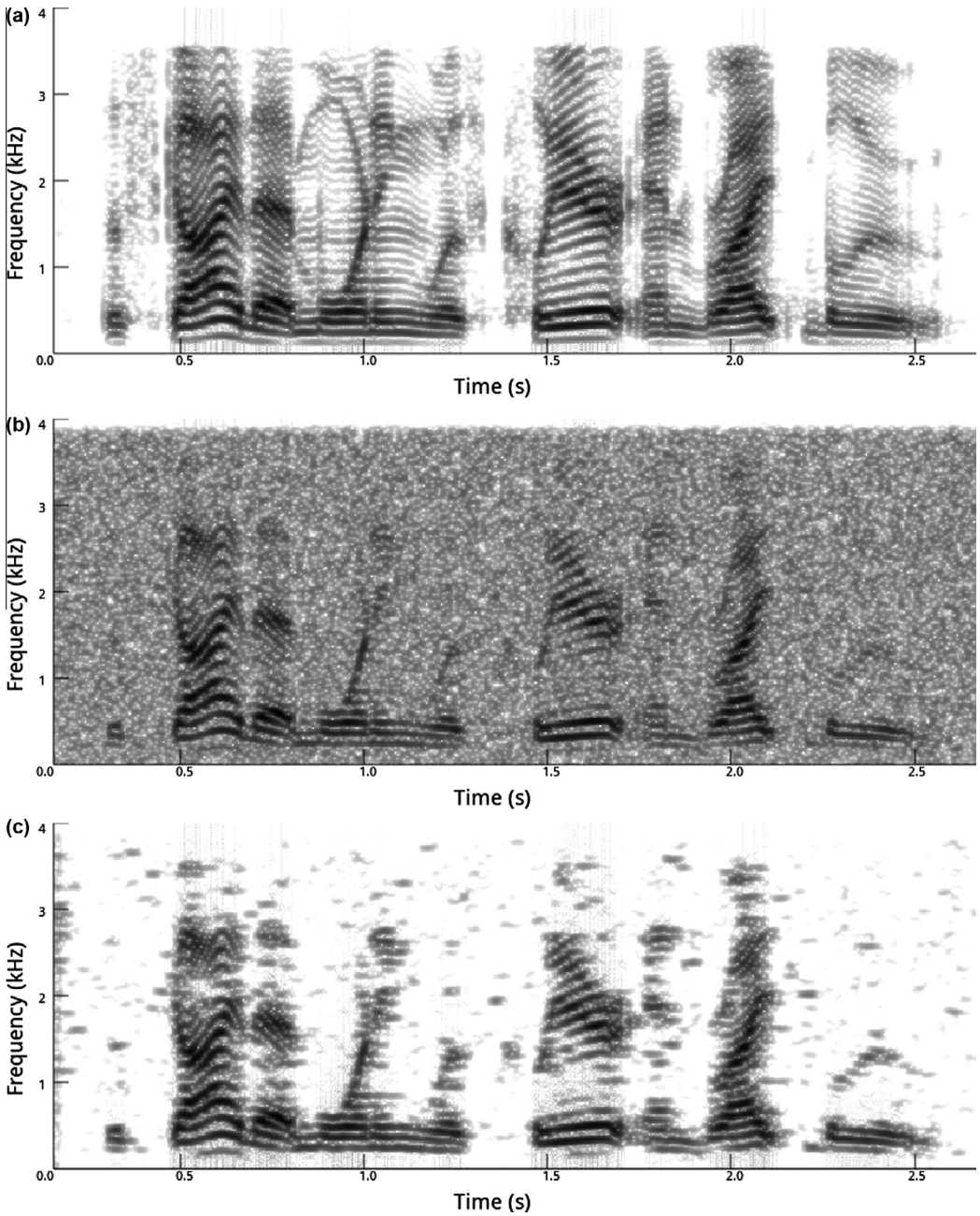


Fig. 17. SE method of speech enhancement; (a) clean speech, (b) speech corrupted with white noise at 15 dB and (c) enhanced speech.

## References


Acero, A., Huang, X., Hon, H., 2001. Spoken Language Processing. Prentice Hall Upper Saddle River, New Jersey.

Allen, I.E., Seaman, C.A. 2007. Likert scales and data analyses. Statistics Round Table, Quality Progress, pp. 64–65.

Alsteris, L., Paliwal, K. 2004. Importance of window shape for phase-only reconstruction of speech. In: Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process (ICASSP), vol. 1, May 2004, Montreal, Quebec, Canada, pp. 573–576.

Alsteris, L., Paliwal, K., 2006. Further intelligibility results from human listening tests using the short-time phase spectrum. Speech Communication 48 (6), 727–736.

ARM, 2009, ARM Ltd., White paper. The arm cortex-a9 processors.

Armstrong, R., 1987. The midpoint on a five-point likert-type scale. Perceptual and Motor Skills 64, 359–362.

Berouti, M., Schwartz, R., Makhoul, J. 1979. Enhancement of speech corrupted by acoustic noise. In: Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process (ICASSP), vol. 4, April 1979, Washington, DC, USA, pp. 208–211.

Billings, D.M., Skiba, D.J., Connors, H.R., 2005. Best practices in web-based courses: generational differences across undergraduate and graduate nursing students. Journal of Professional Nursing 21 (2), 126–133.

Caillaud, B., Cohen, D., 2000. Inter-generational transfers and common values in a society. European Economic Review 44 (4–6), 1091–1103.

Chassaing, R., 2002. DSP Applications Using C and the TMS320C6x DSK. John Wiley & Sons, Inc., New York, NY, USA.

Crochiere, R., 1981. Subband coding. Bell System Technical Journal 60 (7), 1633–1653.

Deci, E.L., Ryan, R.M., 1985. Intrinsic Motivation and Self-Determination in Human Behavior. Plenum Press, New York, USA.

Deci, E.L., Koestner, R., Ryan, R.M., 2001. Extrinsic rewards and intrinsic motivation in education: reconsidered once again. Review of Educational Research, American Educational Research Association (AERA) 71 (1), 1–21.

Donath, J., Boyd, D., 2004. Public displays of connection. BT Technology Journal 22 (4), 71.

Ellison, N.B., Steinfield, C., Lampe, C., 2007. The benefits of facebook friends: exploring the relationship between college students use of online social networks and social capital. Journal of Computer-mediated Communication 12 (4), 1143–1168.

Ephraim, Y., Malah, D., 1984. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. IEEE Transactions on Acoustics Speech and Signal Processing ASSP-32 (6), 1109–1121.

Ephraim, Y., Malah, D., 1985. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. IEEE Transactions on Acoustics Speech and Signal Processing ASSP-33 (2), 443–445.

Finkelstein, J., Wood, J., Cha, E., 2010. Introducing a blackberry e-learning platform for interactive hypertension education. In: Second International Conference on Mobile, Hybrid, and On-Line Learning, ELML '10, February 2010, pp. 77–81.

Freeman, D., Cosier, G., Southcott, C., Boyd, I. 1989. The voice activity detector for the pan-european digital cellular mobile telephone service. In: IEEE Int. Conf. Acoust. Speech, Signal Process. pp. 369–372.

Gartner, September 2011, Gartner Inc. a. Forecast: media tablets by operating system, worldwide, 2008–2015, 3q11 update.

Gartner, November 2011, Gartner Inc.b. Market share: mobile communication devices by region and country, 3q11.

Gonvalves, F., Canesin, C. 2001. Java applets for a www-html-based course in power electronics. In: IEEE 32nd Annual Power Electronics Specialists Conference, PESC, vol. 1, pp. 85–90.

Hassanlou, K., Fathian, M., Akhavan, P., Azari, A., 2009. Information technology policy trends in the world. Technology in Society 31 (2), 125–132.

Hayes, M.H., 1996. Statistical Digital Signal Processing and Modeling. John Wiley & Sons, Inc., New York, NY, USA.

Haykin, S., 1991. Adaptive filter theory. In: Information and System Sciences Series, vol. 2. Prentice Hall, Inc., Englewood Cliffs, New Jersey, USA.

Hermann von Helmholtz, H., 1954. On the sensations of tone as a physiological basis for the theory of music, English ed. Longmans, Green, London, Dover, New York.

Hu, Y., Loizou, P. 2006. Subjective comparison of speech enhancement algorithms. In: Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process (ICASSP), Toulouse, France, May 2006, pp. 153–156.

Hu, Y., Loizou, P., 2007. Subjective comparison of speech enhancement algorithms. Speech Communication 49, 588–601.

Huang, X., Acero, A., Hon, H., 2001. Spoken Language Processing: A Guide to Theory, Algorithm, and System Development. Prentice Hall Upper Saddle River, New Jersey.

Humphreys, L., 2010. Mobile social networks and urban public space. New Media & Society 12 (5), 763–778.

Jackson, J., Barnwell, T., Williams, D., Hayes, M.I., Anderson, D., Schafer, R., 2001. Dsp for practicing engineers: an online course for continuing dsp education. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP '01), vol. 5. pp. 2721–2724.

Joiner, R., Nethercott, J., Hull, R., Reid, J., 2006. Designing educational experiences using ubiquitous technology. Computers in Human Behavior 22 (1), 67–76.

Kay, S., 1993. Fundamentals of statistical signal processing: estimation theory. In: Signal Processing Series. PTS Prentice Hall, Eaglewood Cliffs, New Jersy, USA.

Ko, Y., Duman, T., Spanias, A. 2003. On-line laboratory for communication systems using j-dsp. In: Proceedings of 33rd Annual Frontiers in Education, FIE, November 2003, vol. 1, pp. T3E-13–T3E-18.

Kong, J.S.-L., Kwok, R.C.-W., Fang, Y., 2012. The effects of peer intrinsic and extrinsic motivation on mmog game-based collaborative learning. Information and Management 49 (1), 1–9.

Liu, J., Spanias, A., Banavar, M., Thiagarajan, J., Ramamurthy, K., Hu, S., Zhang, X. 2011. Work in progress x2014; interactive signal-processing labs and simulations on ios devices. In: Frontiers in Education Conference (FIE), October 2011, pp. F2G-1–F2G-2.

Loizou, P., 2007. Speech Enhancement: Theory and Practice. Taylor and Francis, Boca Raton, FL.

Lubis, M.A., Yunus, M.M., Embi, M.A., 2010. Ict and systematic steps in teaching and learning language in the classroom. Procedia – Social and Behavioral Sciences 9, 1055–1061.

Maiti, A., Maiti, A., Maiti, C. 2011. Online technology cad laboratory for microelectronics education. In: 2011 IEEE EUROCON – International Conference on Computer as a Tool (EUROCON), April 2011, pp. 1–4.

Makhoul, J. 1975. Linear prediction: a tutorial review. In: Proc. IEEE, April 1975, vol. 63, no. 4, pp. 561–580.

Martin, R. 1994. Spectral subtraction based on minimum statistics. In: Proc. EURASIP European Signal Process. Conf. (EUSIPCO). Edinburgh, Scotland, September 1994, pp. 1182–1185.

Martin, S., Diaz, G., Sancristobal, E., Gil, R., Castro, M., Peire, J., 2011. New technology trends in education: seven years of forecasts and convergence. Computers and Education 57 (3), 1893–1906.

Milanesi, C., September 2011, Gartner Inc. ipad and beyond: the future of the tablet market.

Milner, B. 2002. A comparison of front-end configurations for robust speech recognition. In: International Conference on Acoustics, Speech and Signal Processing, vol. 1, pp. 797–800.

Mitra, S., 2006. In: Digital Signal Processing: A Computer-Based Approach, vol. 3. McGraw-Hill, Boston, MA, USA.

Nussbaumer, H., 1981. In: Fast Fourier Transform and Convolution Algorithms, 2. Springer-Verlag, Heidelberg, Germany.

NVIDIA, 2010, NVIDIA Corporation, White paper. The benefits of multiple cpu cores in mobile devices.

Ohm, G., 1843. Über die Definition des Tones, nebst daran geknüfter Theorie der Sirene and ähnlicher tonbildender Vorichtungen. Annals of Physical Chemistry 59, 513–565.

O'Neil Jr, H., Perez, R., 2003. Technology Applications in Education: A Learning View. Lawrence Erlbaum Associates, Inc., USA.

Oppenheim, A.V., Lim, J.S. 1981. The importance of phase in signals. In: Proc. IEEE, May 1981, vol. 69, no. 5, pp. 529–541.

Oppenheim, A.V., Lim, J.S., Kopec, G., Pohlig, S.C. 1979. Phase in speech and pictures. In: Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process (ICASSP), April 1979, Washington, DC, USA, pp. 632–637.

Paliwal, K., Alsteris, L. 2003. Usefulness of phase spectrum in human speech perception. In: Proc. ISCA European Conf. Speech Commun. and Technology (EUROSPEECH), September 2003, Geneva, Switzerland, pp. 2117–2120.

Picone, J. 1993. Signal modeling techniques in speech recognition, In: Proc. IEEE, September 1993, vol. 81, no. 9, pp. 1215–1247.

Potts, J., Moore, N., Sukittanon, S., 2011. Developing mobile learning applications for electrical engineering courses. In: Proceedings of IEEE Southeastcon 2011, March 2011, pp. 293–296.

Price, S., Falcao, T.P., 2011. Where the attention is: discovery learning in novel tangible environments. Interacting with Computers 23 (5), 499–512.

Price, S., Rogers, Y., Scaife, M., Stanton, D., Neale, H., 2003. Using tangibles to promote novel forms of playful learning. Interacting with Computers 15 (2), 169–185.

Proakis, J., Manolakis, D., 2007. Digital Signal Processing: Principals, Algorithms, and Applications. Pearson Prentice Hall, New Jersey, USA.

Quatieri, T., 2002. Discrete-Time Speech Signal Processing: Principles and Practice. Prentice Hall, Upper Saddle River, NJ.

Ranganath, S., Thiagarajan, J.J., Ramamurthy, K.N., Hu, S., Banavar, M.K., Spanias, A.S. 2012. Undergraduate signal processing laboratories for the android operating system. In: ASEE Annual Conference, June 2012.

Rienties, B., Tempelaar, D., den Bossche, P.V., Gijselaers, W., Segers, M., 2009. The role of academic motivation in computer-supported collaborative learning. Computers in Human Behavior 25 (6), 1195–1206.

Rosenbaum, E., Rochford, J.A., 2008. Generational patterns in academic performance: the variable effects of attitudes and social capital. Social Science Research 37 (1), 350–372.

Rosenberg, M. 1976. Automatic speaker verification: a review, vol. 64, no. 4, pp. 475–487.Please provide journal title for the reference 'Rosenberg (1976)'.

Salajan, F.D., Schonwetter, D.J., Cleghorn, B.M., 2010. Student and faculty inter-generational digital divide: fact or fiction? Computers and Education 55 (3), 1393–1403.

Shi, G., Shanechi, M., Aarabi, P., 2006. On the importance of phase in human speech recognition. IEEE Transactions on Audio Speech and Language Processing 14 (5), 1867–1874.

Sims, R., 1997. Interactivity: a forgotten art? Computers in Human Behavior 13 (2), 157–180.

Sohn, J., Kim, N.S., Sung, W., 1999. A statistical model-based voice activity detection. IEEE Signal Processing Letters 6 (1), 1–3.

Soliman, S., Srinath, M., 1990. Continuous and Discrete Signals and Systems. In: Information and System Sciences Series, vol. 2. Peason Prentice Hall, New Jersey, USA.

Soloway, E., Norris, C., Blumenfeld, P., Fishman, B., Krajcik, J., Marx, B., 2001. Handheld devices are ready-at hand. In: Communications of the ACM, vol. 44, pp. 15–20.

Song, J., Drennan, J.C., Andrews, L.M., 2012. Exploring regional differences in chinese consumer acceptance of new mobile technology: a qualitative study. Australasian Marketing Journal (AMJ). 20 (1), 80–88.

Srinivasan, K., Gersho, A. 1993. Voice activity detection for cellular networks. In: IEEE Speech Coding, Workshop, pp. 85–86.

Stark, A., Paliwal, K., 2011. Use of speech presence uncertainity with mmse spectral energy estimation for robust speech recognition. Speech Communication 53 (1), 51–61.

Talor, F., 1983. Digital Filter Design Handbook. Marcel Dekker.

Teng, C.-C., Helps, R., 2010. Mobile application development: essential new directions for it. In: Seventh International Conference on Information Technology: New Generations (ITNG), April 2010, pp. 471–475.

Virag, N., 1999. Single channel speech enhancement based on masking properties of the human auditory system. IEEE Transactions on Speech and Audio Processing 7 (2), 126–137.

Wang, D., Lim, J., 1982. The unimportance of phase in speech enhancement. IEEE Transactions on Acoustics Speech and Signal Processing ASSP-30 (4), 679–681.

Wellman, B., 2002. Physical place and cyberplace: the rise of personalized networking. International Journal of Urban and Regional Research. 25 (2), 227–252.

White, Mary-Alice, 1989. Current trends in education and technology as signs to the future. Education and Computing 5 (1–2), 3–10.

Wójcicki, K., Paliwal, K., 2007. Importance of the dynamic range of an analysis window function for phase-only and magnitude-only reconstruction of speech. In: Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process. (ICASSP), Honolulu, Hawaii, USA, April 2007. vol. 4, pp. 729–732.

Yerushalmy, M., Ben-Zaken, O. 2004. Mobile phones in education: a case of mathematics. In: Institute for Alternatives in Education, University of Haifa. Tech. report.