Real Time Determination of an Information Content of an Audio Signal a Multidomain Approach

Engr. Olawore Jelili O.

Abstract: Ability to examine a signal in time, frequency and time-frequency domain gives some statistical features about such signal .It provides valuable insight into the properties that are difficult to discern in a particular domain and the cause for poor and bad performance of such signal. Acoustic analysis of a speech signal in time, frequency and time-frequency is explained briefly. Real time determination of an information content of an audio signal using time, frequency and time-frequency was implemented using Simulink 2016.A multidomain comparison of three adult male's speech was investigated and implemented based on their time, frequency and time frequency representation. Result indicates every individual has a unique way of speaking, some with higher intensity or lower intensity as shown in their time domain representations. There are high level of similarities to all individuals in both their frequency and spectrogram structure display.

Keywords: *EKG, Seismographer, Spectrograph, Spectrogram Spectrum analyzer and Speech enhancement.*

I. INTRODUCTION

Most signals that exist in the real world are captured using sensor, such sensory data ranges from sound waves, seismic vibrations and visual images [1]. Digital Signal Processing (DSP) processes such data, by devising algorithm and techniques to manipulate these signals after they have being converted to digital form [1]. Such manipulation ranges from removing noise from a signal, amplifying a signal, changing a signal to emphasize certain characteristics, for instance adding bass boost to an audio signal [2], image enhancement, speech recognition and generation, data compression for storage and transmission [1].

A signal is an information encoded in a sort of waveform .The need to know the information content of a signal is very important, in order to determine the statistical characteristics of such signal and can serve as a preprocessor to further signal analysis such as speech coding, speech enhancement, sound event detection, pattern recognition e.t.c To analyze an audio signal there is a need to represent such signal in a particular domain(the different ways we can use to represent a signal), which can either be time, frequency or time-frequency domain , wavelet, cepstral, or other domains[3] . In this paper, we are going to focus on time, frequency and time-frequency domain

Time domain is the analysis of physical signal, mathematical functions or environmental data with respect to time. In the time domain, the signal or function's value is known for all real numbers, in continuous time, or at various intervals in the case of discrete time. In time domain, the behavior of the signal is shown over time, the amplitude versus time characteristics of the signal is displayed, and the amplitude might be voltage. It's the only natural way a signal can be represented [4]. Different sources of signals and interference are often illustrated in the time domain [5]. Looking at the amplitude of a signal can yield useful information. For example, looking at the amplitude of a speech signal can give information about the loudness of such signal and the amplitude of a record of a seismographer can provide useful indication of the severity of an earthquake [6].

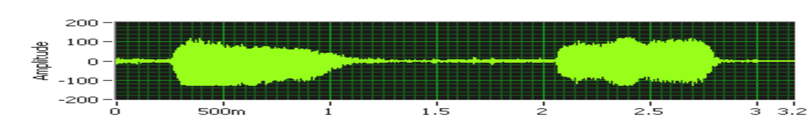


Fig. 1.0: A time domain representation of a speech signal [4].

Figure 1.0 shows how the sound-pressure of a speech signal changes over time. In speech enhancement, the notion is to determine the characteristics of the noise by estimating it during silence periods when no speech is present, the period from 1.4s to 2.0s.The noise computed during the silence period is then removed from the speech signal to

ISSN 2229-5518

obtain a pure speech signal [4].From this waveform as shown in fig. 1.0, some time domain features like mean, root mean square(RMS), standard deviation, variance, zero crossing rate, energy and entropy of energy can be computed which can be used in speech recognition and speech enhancement. An oscilloscope is an instrument use for displaying and analyzing a signal in time domain. The vertical axis of the display screen represents voltage, and the horizontal axis represents time. The oscilloscope can measure the amplitude, period, and frequency of a signal and also, the duty cycle, pulse width, rise time, and fall time of a pulse waveform.

Frequency domain refers to the analysis of signals or mathematical functions with respect to frequency. It shows how the signal's energy lies over a range of frequencies [8].It is very important in the field of signal processing. It is applied in various areas such as communications, geology, remote sensing, and image processing e.t.c [8]. A frequency-domain representation has information on the phase shift that must be applied to each sinusoid or frequency component, for recovering the original time signal, by combining the individual frequency components. A mathematical function or signal can be converted between the time and frequency domains, by using a pair of mathematical operators called a transform. An example is the Fourier transform and also use an inverse Fourier transform to convert it back to time domain signal or function.

The unique characteristic of the frequency domain is that it is not real; it's a mathematical construct on a raw time domain signal (the only reality). The unique rule in the frequency domain is that the only kinds of waveforms that exist are sine waves. They are basically the language of the frequency domain [7]. However, some problems are easily analyzed in the frequency domain than in the time domain. For instance, the concept of bandwidth is inherently a frequency-domain idea or the total frequency spectrum of speech is used to determine the bandwidth allocated to a communication channel for the transmission of speech signal. When we have multiple frequency values; the collection of amplitudes is called the spectrum [7].Some frequency domain methods are Fourier series, Fourier transform, Laplace transform, Z transform and Wavelet transform. Fourier series is applied to repetitive signals and oscillating systems. Fourier transform is applied to nonrepetitive signals and transients. Laplace transform is applied to electronic circuits and control systems. Z transform is applied to discrete-time signals and digital signal processing. Wavelet transform is applied to image analysis and data compression. In linear system analysis, frequency domain representations are used often [5].System behavior and signal transformations are more convenient when analysis is done in the frequency domain [5].

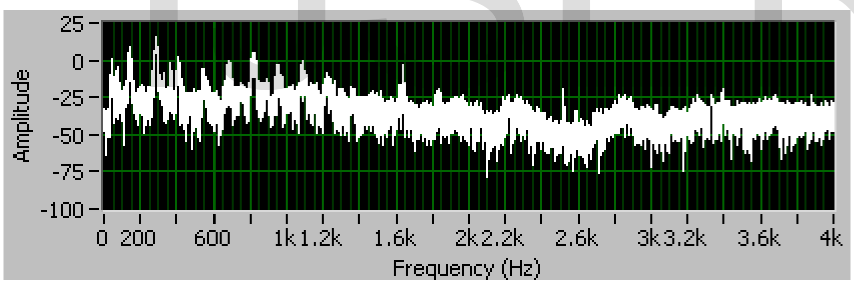


Fig. 1.1: A frequency domain representation of a speech signal[5].

In acoustic analysis of a speech signal, it is vital to identify the spectral content of the signal. Speech signal can be easily analyzed in frequency domain than in time domain [9]. The human ear focuses more on some aspects of the speech than its phase or timing aspects [9]. Spectral analysis is used to extract most of the parameters of speech signal [9]. The frequency characteristics of a speech signal can be determined by taking an FFT of the time waveform [7]. The square of the magnitude of the FFT is known as power spectrum. It shows how the energy of a signal is distributed in the frequency domain [7].From the waveform as shown in fig. 1.1, some frequency domain features like spectral centroid and spread, spectral entropy, spectral flux, spectral rolloff, MFCC and chroma vector can be computed which can be used in speech recognition and speech enhancement. A spectrum analyzer is a device that shows the frequency components of a signal otherwise referred to as spectrum. By analyzing the spectral content of the signal, distortion, power, dominant frequency, bandwidth , harmonics and other spectral components of a signal can be observed that are not easily discernible in time domain waveforms. A spectrum analyzer helps in visualizing, analyzing and real-time measure of the frequency domain feature of a signal [10].The spectrum analyzer software in MATLAB displays and compute the spectrum, power density function and spectrogram of various signal such as biosignals, audio, communication signals, speech etc.. The spectrum analyzer software can measure the harmonic distortion such as SNR,SINAD,SFDR, third order intermodulation distortion(TOI),peak finder, spectrogram mode, peak to –average power ratio, adjacent channel power ratio(ACPR) and complementary cumulative distribution function(CCDF)[10].

In real life scenario, many signals are nonstationary for instance speech signal, music and images, their spectrum changes over time. Time-frequency analyses are applied to nonstationary signals. It provides both temporal and spectral information. It answers questions such as: How do I visualize the time-frequency information of my signal? When is a particular frequency component present in my signal? How do I increase time or frequency resolution? Techniques such as the Short-time Fourier Transform and the Wavelet Transform use the concept of time-frequency representations of a signal [11]. Time-frequency analysis is mostly done by segmenting a signal into short intervals and estimating the spectrum over sliding windows [12]. An audio signal is broken down into segments say 2 to 40 milliseconds, and the fast Fourier transform computes the frequency spectrum of each segment. These spectra are placed side-by-side, and converted into an image, with the low amplitude, light (least energy) and high amplitude, dark grayscale image (most energy) [1]. Moments of silence, and frequency regions with little energy, appears white, dark regions shows area of most energy which might be caused by formant vibration in a speech signal, vocal fold closures or harmonics [13]. Fig. 1.2 shows a spectrogram display of a speech signal. A visual display of the time-frequency representation of an acoustic signal using STFT is known as spectrogram [14].

A spectrogram displays a three-dimensional spectral representation of speech utterances typically about 2–3 s in duration. It translates the speech signal into a visual domain; the energy content of a signal is expressed as a function of frequency and time. The vertical axis indicates frequency, the horizontal axis is time, and amplitude is display on a grey-scale, with the power at any instant denoted by a color [13].Two types of spectrogram is used, the wideband spectrogram and narrow band spectrogram. The wideband spectrograms display individual pitch periods as vertical striations corresponding to the large speech amplitude each time the vocal cords vibrate. Voicing can be easily detected visually by the presence of these periodically spaced striations; it has a detailed time resolution and typically use a window about 3 ms long with a bandwidth of 300 - 500Hz. The narrowband spectrograms use a long time-span window to represent the harmonic structure of speech instead of pitch periods, with fine frequency resolution, at the expense of poor time resolution. It often uses a 20-ms window with a bandwidth 45- 50Hz.

The length of the segment is chosen as a tradeoff between better frequency resolution (longer segments) and better time resolution (shorter segments) [1] [12]. Segmenting a speech signal accurately and determining periods in voiced speech, time resolution is very useful while different sounds can be identified with a good frequency resolution.

This short-time frequency analysis of a continuous audio signal and its representation as a two-dimensional time-frequency image visually is referred to as spectrogram [15].Spectrogram has being used to distinguish different sound event in unstructured environments as reported in [15]. A sound spectrograph is an instrument that displays a spectrogram [16]. The spectrogram analyses vocal output, identify the strength and frequencies of formants, and for real-time feedback in voice therapy and training. Other time-frequency domain techniques are Gabor spectrogram, Wavelet transform, Wigner distribution and Cohen class transforms.

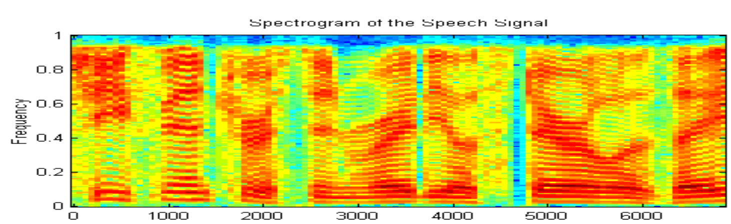


Fig. 1.2: Spectrogram of a speech signal [10].

II RELATED WORK

Ye Wang and Mikka Vilermo applied time-frequency signal processing method and representation to audio denoising and auditory modeling[18]. P. R. Ghosh et al., used Hilbert spectral analysis to represent information content of an audio signal and showed that Hilbert spectral analysis approach is better than Fourier based method[19]. K. M. Hassan et al. also used an enhanced time-frequency representation of an audio signal known as EMD-2TEMD based approach for source separation technique [20].

III METHODOLOGY

In this paper, Simulink a graphic-based programming, which uses block diagram environment for model-based design and multidomain simulation, was used to determine the information content of an audio signal. The blocksets are all selected from the Simulink library browsers. The setup is connected as shown in Fig. 2.0, the Audio Device Reader blockset records audio stream from the computer audio device in real time sampled at 44100Hz. The other three blockset where connected to the audio device reader blockset. The first blockset named spectrum analyzer1 is a spectrum analyzer but configured to display the spectrogram, which depicts a spectrograph. It displays the spectrogram of the audio signal as it is being recorded. The second blockset, time scope displays the time domain waveform of the audio signal as it is being recorded. It also depicts a real life oscilloscope. The third blockset, a spectrum analyzer displays the spectral content of the audio signal as it is being recorded. All connected blocksets operate and display in real-time. When the simulation starts, a 10 seconds audio signal was recorded and all blocksets displayed the time, frequency and time-frequency waveform as the sentence was spoken. The following sentence was spoken," The knowledge of the range of speech and noise intensity levels in real world scenarios are critical term to the design of speech enhancement algorithm".

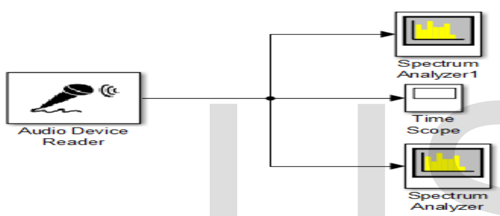
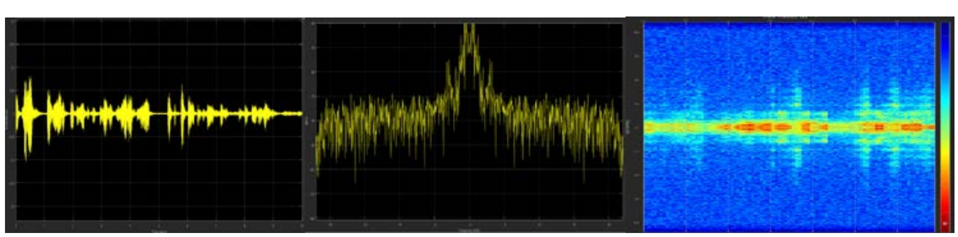


Fig. 3.0: Real time determination of an information content of an audio signal

IV RESULTS

A multidomain approach (time, frequency and time-frequency) to determining the information content of a speech signal was implemented. Speech signals of three adult, males was analyzed in order to compare their time, frequency and time-frequency characteristic or information content. The male ages are thirty-seven, thirty two and thirty-four. All the three individuals spoke for 10 seconds and read the same sentence. Fig. 4.0 (a) shows the time domain information of the speech signal from the first male(37) , it shows a maximum peak which represent the highest amplitude or pitch of the word spoken of 8.173-02 at 0.520s,a minimum peak of – 1.061e-01 at 0.342s, a peak to peak of 1.879e-01, mean of -1.222e-04,median of -6.104e-05 and RMS value of 1.048e-02. Fig. 4.0(b) shows the relative intensity of the energy of the signal at each frequency for the entire speech signal. The logarithm of the magnitude of the FFT (amplitude) on the y-axis with range from -120 to -40dBm and the frequency range in kHz at the horizontal axis with range from -22 to 22 kHz. The frequency domain representation shows some peaks at -34.7638 at -0.546 KHz, 34.7638 at 0.546 KHz and -34.8875 at -0.373 KHz. In its spectrogram display there is higher energy concentration from -1 to 1 KHz at -55dBm, with energy spreading lightly from -10 to 10 KHz with an exception of -1 to 1 KHz.



ISSN 2229-5518

Fig. 4.0: (a) Time domain waveform of male speech signal (37) (b) The spectrum display of a speech signal (37) (c) A spectrogram of a speech signal (37).

The time domain information of the speech signal from the second male (32) from figure 4.1(a), It shows a maximum peak of 2.960e-02 at 4.337s,a minimum peak of – 3.333e-02 at 4.702s, a peak to peak of 6.293e-02, mean of -1.298e-04,a median of -6.104e-05 and RMS value of 5.119e-03. Fig. 4.1(b) shows the relative intensity of the energy of the signal at each frequency for the entire speech signal. The logarithm of the magnitude of the FFT (amplitude) on the y-axis with range from -120 to -40dBm and the frequency range in kHz at the horizontal axis with range from -22 to 22 kHz. The frequency domain representation shows some peaks at -24.6854 at -0.144 KHz, -24.6854 at 0.144 KHz and -26.2043 at -0.402 KHz. In its spectrogram display, The yellow red indicates the volume of the sound (pitch) and the highest energy concentration, display shows highest pitch within -1 to 1 kHz at -55dBm for the entire time range, with energy spreading with little intensity or low pitch from -10 to 10 KHz .

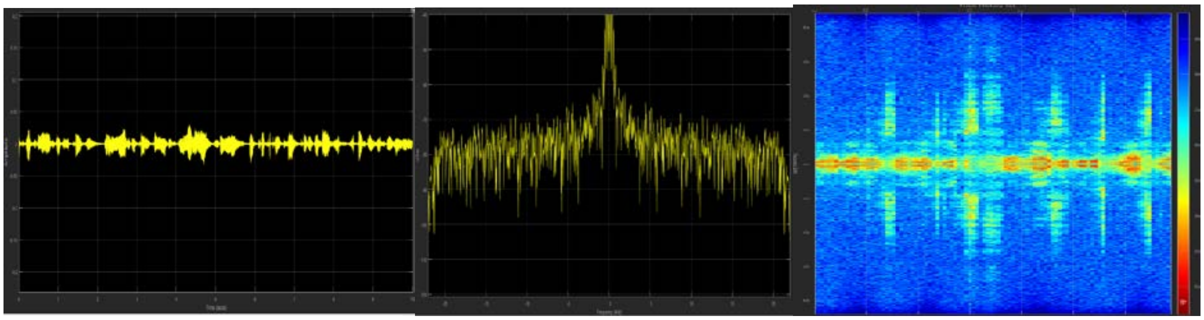


Fig. 4.1: (a) Time domain waveform of male speech signal (32) (b) The spectrum display of a speech signal (32) (c) A spectrogram of a speech signal (32).

The time domain information of the speech signal from the third male (34) from figure 4.2(a) It shows a maximum peak of 4.370e-02 at 0.283s, a minimum peak of – 5.313e-02 at 0.284s, a peak to peak of 9.683e-02, mean of -1.210e-04, a median of 0.000e+000 and RMS value of 6.727-03, a 10 seconds real time recording was done but the speaker already completed the entire sentence around 9.15 seconds, with the remaining seconds indicating moment of silence. Fig. 4.2(b) shows the relative intensity of the energy of the signal at each frequency for the entire speech signal. The logarithm of the magnitude of the FFT (amplitude) on the y-axis with range from -120 to -40dBm and the frequency range in kHz at the horizontal axis with range from -22 to 22 kHz. The frequency domain representation shows some peaks at -37.5236dBm at -0.029 KHz, -37.5236dBm at 0.029 KHz and -47.4343dBm at -0.201 KHz. In its spectrogram display, the yellow red indicates the volume of the sound (pitch) and the highest energy concentration, display shows highest pitch within -1 and 1 kHz at -55dBm for the entire time range, with energy spreading with little intensity or low pitch from -8 to 8 KHz with exception of -1 to 1 kHz.

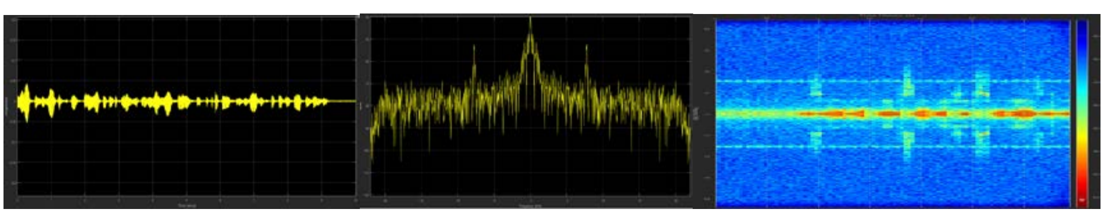


Fig. 4.2: (a) Time domain waveform of male speech signal (34) (b) The spectrum display of a speech signal (34) (c) A spectrogram of a speech signal (34).

.

V DISCUSSION AND CONCLUSION

Visualizing a signal with a plot is all we need to decode the information we are looking for. For instance, a doctor looking at the plot of an electrocardiogram (EKG) can know how well the heart of the patient is working [6]. Fig. 4.0 (a) shows the time domain information of the speech signal from the first male (37), shows a moment of silence from 0.8 to -1.2s, 4.7 to 5.3s and 9.2 to 10s. In its spectrogram display, the deep blue indicates the lowest energy intensity, the light blue indicates lower energy intensity, the yellow indicates a moderate high intensity and red indicates the highest intensity. Figure 4.1(a) also shows the time domain information of the second male (32), speech signal which indicates some moment of silence. In its spectrogram display, the yellow red indicates the volume of the sound (pitch) and the highest energy concentration. The time domain information of the speech signal from the third male (34) from figure 4.2(a), it shows a moment of silence from 9.2 to 10 seconds. In its spectrogram display, the yellow red indicates the volume of the sound (pitch) and the highest energy concentration. Result indicates every individual has a unique way of speaking, some with higher intensity or lower intensity as shown in their time domain representations; there are high level of similarities to all individuals in both their frequency and spectrogram display. In audio signal analysis, asking which domain is more important is equivalent to asking the question which is more important in life, food, air or water. Each domain has its strength and weakness [21]. Having some intuition about a signal in time, frequency or time-frequency domain can give valuable insight to the cause for good or poor performance [21]. Frequency-domain view of a signal shows information about your signal that might not be discernible in the time-domain representation but does not show how the frequency content of a signal changes over time (time-frequency). Some basic questions like I'm I interested in a time domain, frequency domain or time-frequency domain parameter of a signal will give you an intuitive approach to know which domain to use. You have to approach solving a problem based on the domain that produces the optimum result, whether in time, frequency, time-frequency or other domain approach. This work can serve as a preprocessor to other digital signal processing areas such as speech processing, speech recognition, speech enhancement, echo cancellation, environmental sound detection etc..

REFERENCE

1) S.W Smith, "The Breadth and Depth of DSP" *The Scientist and Engineer's Guide to Digital Signal Processing*, 2nd Ed., San Diego, California Technical Publishing 1999, pp.1-5.

2) An Introduction to Signals. [Online] Available: http://www.facstaff.bucknell.edu/mastascu/elessonshtml/Signal/Signal1.htm [Accessed: May. 28, 2017].

3) F. Alías et. al., "A Review of Physical and Perceptual Feature Extraction Techniques for Speech, Music and Environmental Sounds" , Journ. of *Appl. Sci.* 2016,vol *6*(5), Switzerland ,pp142-186.[Online] Available: *www.mdpi.com/2076-3417/6/5/143*. [Accessed: June 16,2017].

4) E. Bogatin, "Time and Frequency Domains", *Signal and Power Integrity*, 2nd Ed., Prentice Hall [Online] Available: http://www.informit.com/articles/article.aspx?p=1374896 [Accessed: May. 28, 2017].

5) LearnEMC (2017), "Time/Frequency Domain Representation of Signals". [Online] Available: https://learnemc.com/time-frequency-domain [Accessed: May. 30, 2017].

6) E. Corona," Basic Signal Analysis", ME231Measurements Laboratory Spring 2003.[Online] Available: http://www3.nd.edu/~pdunn/www.text/ECbasiganal.pdf .[Accessed: June 06, 2017].

7) National Instruments, "Introduction to Time Frequency Analysis (Advanced Signal Processing Toolkit)" [Online] Available: http://zone.ni.com/reference/en-XX/help/372656C-01/lvasptconcepts/tfa_intro/ [Accessed: May. 28, 2017].

8) Mathworks," Practical Introduction to Frequency-Domain Analysis", [Online] Available: https://www.mathworks.com/help/signal/examples/practical-introduction-to-frequency-domain-analysis.html [Accessed: May. 28, 2017].

9) L. Deng and D. O' Shaughnessy,"*Speech Processing A Dynamic and Optimization-Oriented Approach"* Marcel Dekker Inc., New York, NY. , June 2003,pp 5-17.

10) Mathworks," Visualize, measure and analyze multichannel spectrum", [Online] Available: https://de.mathworks.com/discovery/spectrum-analyzer-software.html [Accessed: August. 11, 2017].

11) Tryengineering," What is the difference between time domain and frequency domain analysis? ", [Online] Available: http://tryengineering.org/ask-expert/what-frequency-domain-analysis-what-difference-between-time-domain-and-frequency-domain[Accessed: May. 28, 2017].

12) Mathworks," Practical Introduction to Time-Frequency Domain Analysis", [Online] Available: https://www.mathworks.com/help/signal/examples/practical-introduction-to-time-frequency-analysis.html [Accessed: May. 28, 2017].

13) UCL,"Spectrograms". [Online] Available: http://www.phon.ucl.ac.uk/courses/spsci/acoustics/week1-10.pdf [Accessed: June 06, 2017].

14) R. K. Reddy et al.,"Categorization of environmental sounds ", Journ. of Biol Cybern 2009, pp299–306. DOI 10.1007/s00422-009-0299-4

15) J. W. Dennis," Sound Event Recognition in Unstructured Environments using Spectrogram Image Processing" Ph.D. dissertation, Computer Engr., Nanyang., Singapore, 2014.

16) NCVS," The Spectrograph & Spectral Analysis". [Online] Available: http://www.ncvs.org/ncvs/tutorials/voiceprod/tutorial/spectral.html [Accessed: June 06, 2017].

17) Time and Frequency Representations Overview.[Online] Available: http://www.seas.ucla.edu/dsplab/tfr/over.html [Accessed: May. 30, 2017].

18) Ye Wang and Mikka Vilermo, "Audio Signal Representation and Processing in Time-Frequency Domain" ICMC Proc.1999 pp264-268.

19) P. R. Ghosh et al. ," Analysis of Audio Signals with Noise Assisted Empirical Mode Decomposition" Int. Journ. of Signal Processing, Image Processing and Pattern Recognition vol. 8, no. 4(2015), pp. 73-88.

20) K. M. Hassan et al.," An Improvement in Representation of Audio Signal in Time-Frequency Plane using EMD-2TEMD Based Approach" [Online] Available: www.sersc.org/journals/IJSIP/vol8_no4/7.pdf [Accessed: August. 13, 2017].

21) G.L. Cheminant (2013)," Understanding Measurement in the time And Frequency Domains". [Online] Available: http: //www.electronicdesign.com/test-amp-measurement/understanding-measurement-time-and-frequency-domains.[Accessed: May. 30, 2017].