A Multidomain Characterization of an Acoustic Signal and an Approach towards a Generic Noise Reduction Algorithm

Engr. Jelili O. Olawore, Dr Christopher U Ngene , Dr Abdulfattah A. Aboaba

Abstract— During mobile communication, a speech signal contains information from varieties of sources: speaker, environment sound and transmission channel. Due to effect of background noise otherwise referred to as environmental sounds such as dog barking, siren, car horn, restaurant babble noise, gun-shot, street music, white noise e.t.c, leads to the degradation of speech quality. In order to discriminate between different kind of signals (speech, music and environmental sounds),for an efficient speech enhancement algorithm, pre-processing of such signal is required by having a apriori characteristic of such signals based on their distribution of their energy content over time, frequency and time-frequency domain and also, having the range of noise level encountered in real life scenario. There are uncountable numbers of sounds that interferes with the main speech during mobile communication. In this paper, uncontrollable sounds with high level of noise intensity is considered, to narrow the scope of acoustic signals (environmental sounds) considered for characterization and in an attempt to develop a generic noise reduction algorithm which can perform efficiently in all types of noise environments. In this paper, multidomain characterization of some datasets of uncontrollable environmental sounds was analyzed and presented.

Keywords— *Acoustic, ECG, EEG, EMG, multidomain, Speech enhancement and Spectrogram*.

I. Introduction

Mobile communication is the mostly commonly used medium for speech communication device. Its versatility to be used anywhere and everywhere leads to the degradation in the speech quality due to the effect of background noises. For an efficient speech enhancement algorithm to be developed, characterization of various environment sound or background noises that affects speech during mobile communication has to be studied [6]. Pre-processing of such acoustic signal is required by having an apriori characteristic of such signals based on their distribution of their energy content over time, frequency and time-frequency. The environment sound or acoustic noise could be additive (added to the clean signal), multiplicative or convolutive [6][7] to the speech signal, if the source is from a highly reverberant room. Also, based on their source of generation, the temporal behavior may be stationary or non-stationary [6].The noise might be broadband or narrowband.

Different kinds of acoustic signals (speech, music and environmental sound) have different characteristics depending on the distribution of their energy content over time, frequency and time-frequency [6].It is very difficult or impossible to develop an algorithm that works in all type of noise environment as reported in [9], due to the numerous source of sounds that interferes with speech signal during communication. Uncontrollable sounds with high level of noise intensity are considered to narrow the scope of acoustic signals considered for characterization. Environmental sounds where characterized based on their spectral content as reported in [7], for mobile communication. Environmental sounds where also characterized based on their time-frequency content (spectrogram) as reported in [2], for sound recognition. This paper investigates characterization of acoustic signals with emphasis on uncontrollable environmental sound with high level of noise intensity using a multidomain approach, in the time domain (operating directly on the audio waveform) where parameters like mean, root mean square (RMS), standard deviation, variance value can be estimated [8], in the frequency domain (after a spectral transformation of the audio signal) where power spectral density(psd)[8]is estimated and time-frequency domain, a visual display of the acoustic signal using the short-time Fourier transform (STFT) in the time–frequency representation, otherwise known as the spectrogram, it

ISSN 2229-5518

provides a time-frequency representation of a signal. It was reported in [6], for an efficient speech enhancement algorithm to be developed, it must be able to work within -5 to 15db SNR.

II. LITERATURE REVIEW

Environmental sounds where characterized based on their spectral content as reported in [7],for mobile communication.Environmental sounds where also characterized based on their time-frequency content (spectrogram) as reported in [2],for sound recognition. Most energy in human speech is concentrated in the frequency band between 500 and 2000Hz [1]. It is indentified by the pauses between the words as shown in figure 2.0(a) .The spectrogram of a hi-fi version of a speech signal shows most of the energy is concentrated below 4KHz as shown in figure 2.0(b). Speech has a narrow bandwidth, with much of the energy concentrated in the lower frequencies [2].

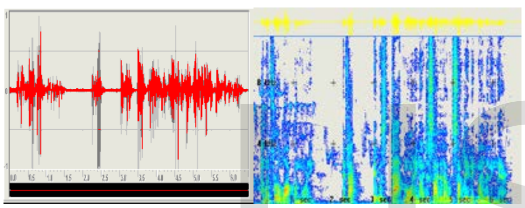


Fig. 2.0: (a) Time domain waveform of a speech signals (b) The spectrogram of a hi-fi version of a speech signal [1].

Music covers a larger frequency range than human voice with more energy concentrated at lower frequency but higher frequencies are used to differentiate different musical instruments [1]. There is little occurrence of pauses between words or sentences when compared to speech as shown in figure 2.1. Spectrogram of a music signal with frequency within the range of 50Hz – 20000Hz is shown in figure 2.1(b). Music has a broader bandwidth when compare to speech, containing energy up to 20 kHz and above [2].

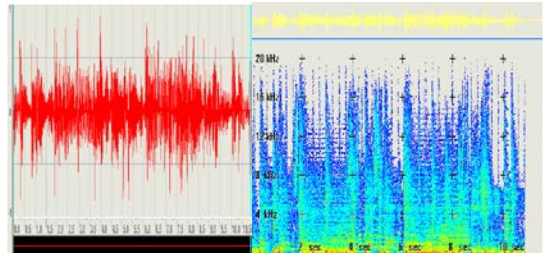


Fig. 2.1: (a) Time domain waveform of a Musical signal (b) The spectrogram of a musical signal.

Environmental Sounds: It is very challenging to enumerate the general characteristics of environment sounds when compare with music or speech, due to different types of sources and environment that it occurs [2].Environmental sounds are unstructured due to their variety of sound source, in which both speech and music posses both formantic and harmonic structures [4]. Environmental sound can be further subdivided into stationary and nonstationary noise based on their source of generation. In stationary noise, the spectral content does not change as a function of time, space, or some other independent variable while nonstationary the spectral content changes over time, space, or some other independent variable [5]. Examples of stationary sound/noise are siren, wind, white noise, colored noise, raining dropping on a roof, generator noise, jet engine noise, fans, air-conditioner, car noise (relatively) and sinusoidal signal. Examples of nonstationary noise traffic, seismic signal, ECG, EEG, and EMG.

Sounds can be generally classified as depicted in figure 2.2

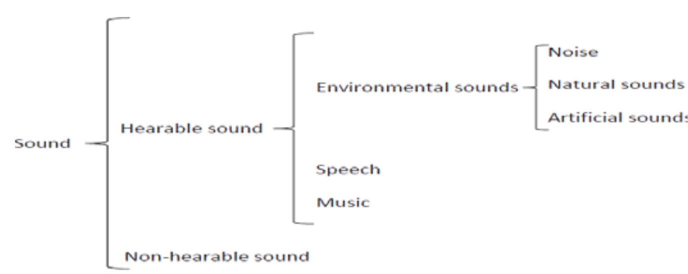


Fig. 2.2: A generic sound classification system [3].

Table 2.1: Acoustical characteristics of speech, music and environment sounds [2] [3].

Acoustical Characteristics	Speech	Music	Environment Sounds
No. of Classes	No. of Phonemes	No. of Tones	Undefined
Window Length	Short (fixed)	Long(fixed)	Undefined
Bandwidth	Narrow	Broad	Broad Narrow
Harmonics	Clear	Clear	Clear Unclear
Repetitive Structure	Weak	Weak	Strong, Weak
Source	Human speech Production System	Instrument or Human Speech production System	Any other audio source
Temporal and Spectral Characteristics	Short durations (40–200ms), A lot of harmonic content around 500 Hz to 2 KHz with some noise-like sounds	From short to long durations (40–1200ms), with a mix of steady and transient sounds organized in periodic structures, A lot of harmonic content in the full 20 Hz to 20 KHz audio band and some inharmonic components.	From short to very long durations (40–3000ms), with wide range of steady and transient type of sounds, and also a broad range of harmonic and inharmonic content

III. METHODOLOGY

Uncontrollable environmental sounds with high intensity level that might interfere with the main speech during mobile communication were considered for characterization. Datasets of thirty-three (33) classes were investigated for characterization. They were collected from various sources, urbansound data sets [10], Dan Ellis sound examples [11], signal processing information base noise data [12], environmental sounds samples recorded through a TECNO-C9 mobile phone which were all .aac format and converted to .wav format using an online converter at [13], the acceptable format for sound signal analysis in MATLAB and some other random sources. All original sampling frequency or rate where maintain to avoid introduction of buzz sounds due to aliasing effect [14].Original sampling frequency of 19.98KHz,44.1KHz,48KHz and 96KHz were all maintained. The datasets include horn, factory noise(electrical welding equipment),crowd clapping, crowd clapping and hissing, children playing, baby crying, dog barking, drilling, gunshot, siren, street music with multiple noises, babble(restaurant), glass breaking, park, market place, train(moving), train(announcement), train(arriving), generator noise(white noise),cockpit noise, moving tricycle, rainfall, thunder, parade drum, trumpet, church drum, party music, street noise (traffic, passing vehicle with music on, horning),call for prayer, lawn mower, helicopter starting, church drum and church singing . For illustration purpose five of the datasets will be considered for analysis which includes cockpit, factory noise, generator noise, babble and horn sound. Each recording were considered for analysis. A digital signal processing tool in MATLAB, sptool was used for both time and frequency domain analysis but can be otherwise done at the MATLAB command window. Spectrogram was typed at the command window of MATLAB for time-frequency domain analysis of each of the signals.

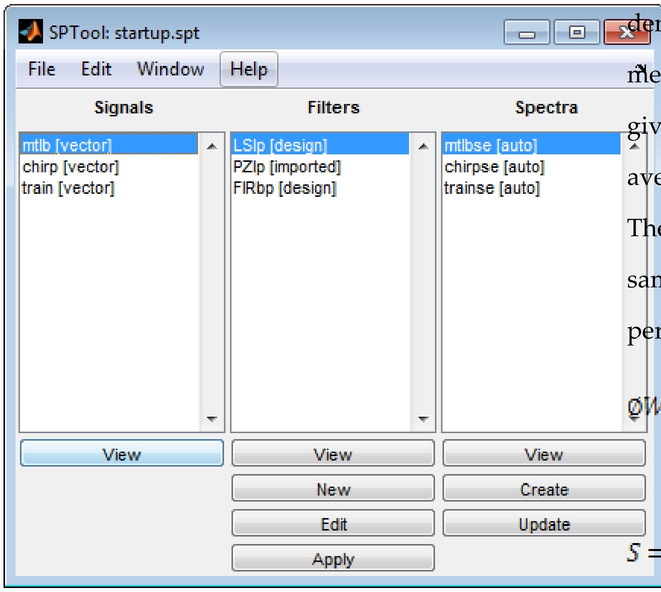


Fig. 3.0: SPTOOL graphical user interface (GUI), adapted from [15].

After signal analysis is done at the MATLAB command window or editor, the analyzed signal can be imported to the sptool window, for more information on how to use sptool,[15] can be explore. Time domain waveform of a signal or more than one signals can be viewed at the signal browser denoted by the word Signals from figure 4.0 above. Statistical values of a signal such as RMS, mean, median, maximum, minimum and peak-to-peak difference values of a selected signal can be viewed in this panel. Frequency domain analysis is done at the spectra panel, various parametric, nonparametric and subspace methods of power spectral density estimation are included. Parametric methods such as Burg method, Modified Covariance, Yule-Walker autoregressive method and nonparametric method such as the FFT method, the multitaper method, the Welch's method, also subspace methods such as MUSIC and eigenvector were all included in sptool. For more information on power spectral density estimation [7],[16] can be explore and in depth explanation may be found in [17].Welch's method of power spectral density (psd) estimation was used in this paper. Welch's method inherently removes spurious spectral peaks and gives more accurate power measurement through averaging as many periodograms [18].

The Welch's PSD for a given sample x (n), having N-1 samples is determined by averaging the windowed periodograms

$$\emptyset W(w) = \frac{1}{S}\sum_{j=1}^{S} \emptyset_j(w) \qquad (1)$$

Where S is the segmented samples

$$S = \frac{N}{M} \qquad (2)$$

Where N is the number of samples and M is the length of data segment. The windowed periodogram in respect to window $y_j(t)$ is given as

$$\emptyset_j(w) = \frac{1}{MP}\left|\sum_{t=1}^{M} V(t)y_j(t)e^{-iwt}\right|^2 \qquad (4)$$

P represents power of the temporal window {V (t)}:

$$P = \frac{1}{M}\sum_{t=1}^{M} |V(t)|^2 \qquad (5)$$

Spectrogram provides a time-frequency representation of a signal. It shows a plot of frequency as a function of time, with the power at any instant denoted by a color. Bright color signifies high intensity of power at that instant. In MATLAB dark red is the largest intensity, dark blue is the lowest while bright yellow or green are in the medium. For more information on the use of spectrogram in MATLAB, help spectrogram should be typed at the command prompt.

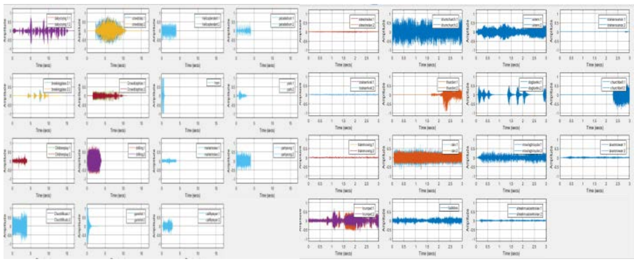


Fig. 3.1: Time domain waveforms of various uncontrollable environment sounds with high level of intensity using MATLAB sptool.

IV. RESULTS

Original sampling frequency where maintain from various sources. Babble noise was acquired from 100 people speaking in a restaurant. It was considered as one of the uncontrollable environmental sounds with an original sampling frequency of 19.98 KHz. Using sptool, the time domain information shows a maximum peak of 2.702e-01 at 2.595,a minimum peak of -3.575e-01 at 2.391s, a peak to peak of 6.277e-0100,mean of 2.014e-03,median of 2.441e-03 and RMS value of 6.199e-02.The power density(PD) increased from -58dB till -55dB at 0Hz and increases thereafter till -50dB (maxPD) and to a minimum PD at -98dB at 10KHz. There are crests with a significant one at -84.41(6340) Hz. The babble spectrogram shows high level of power intensity concentration (energy) at 5 KHz below which his typical of a speech like signal and lower power intensity (energy) from 10 KHz to 20 KHz above.

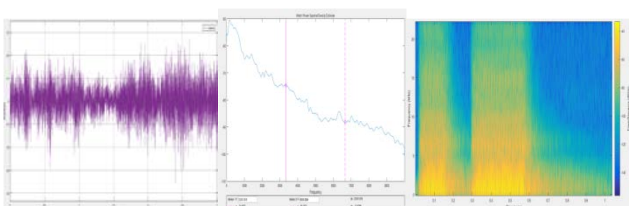
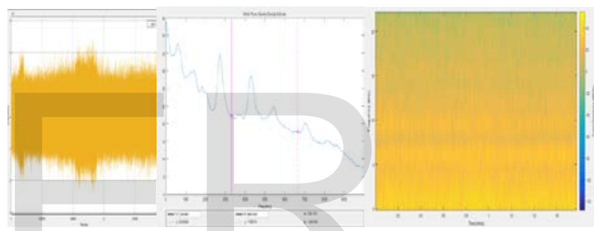


Fig. 4.0: (a) Time domain waveform of babble signal (b) The PSD display of a babble signal (c) A spectrogram of a babble signal.

Cockpit noise was recorded at the co-pilot's seat in a two-seat F-16, traveling at an altitude of 300-600 feet and a speed of 500 knots, with an original sampling frequency of 19.98KHz. Using sptool, the time domain information shows a maximum peak of 2.514e+04 at 2.196m,a minimum peak of -2.440e+04 at 2.215ms, a peak to peak of 4.954e+04, mean of 1.408e+02,median of 1.370e+02 and RMS value of 4.478e+03.The PD increased from 46dB(0Hz) till 50dB at 0Hz and decrease downward till 24dB(2356Hz) and goes upward till 39.53dB at 2740Hz.Thereafter increased upward till 33.47dB(4297Hz) and goes down to 7.7dB at 10KHz and goes down further to a minimum PD 4.71dB at 10KHz. There are noticeable crests at 39.5dB (2740Hz), 33.33dB (4318Hz), 24.80dB (5450Hz) and 19.95dB (7088Hz).It's spectrogram shows a uniform high energy intensity across the plot, with higher energy distribution at 10 KHz below.



. 4.1: (a) Time domain waveform of a cockpit noise (b) The PSD display of a cockpit noise (c) A spectrogram of a cockpit noise.

Factory noise acquired near plate-cutting and electrical welding equipment, with an original sampling frequency of 44.1 KHz. Using sptool, the time domain information shows a maximum peak of 2.368e+04 at 1.108m, a minimum peak of -2.064e+04 at 1.704m, a peak to peak of 4.431e+04,mean of 1.418e+02,median of 1.410+02 and RMS value of 2.713e+03.The PD increases from 40.20dB upward till 43.50dB at 0Hz and thereafter decrease downward till 14.5dB(10KHz).There was further decrement till 11.32dB(minimum PD) at 10 KHz. It was observe there was virtually a uniform PD from 5125Hz till 6886Hz. Its spectrogram shows a partial distribution of high intensity

ISSN 2229-5518

energy across the plot but with higher energy concentration at 5 KHz below.

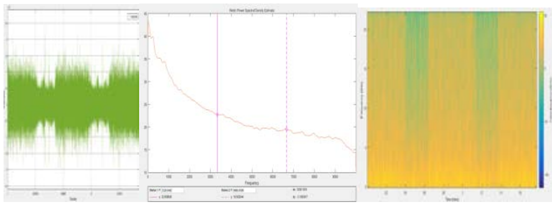


Fig. 4.2: (a) Time domain waveform of a factory noise (b) The PSD display of a factory noise(c) A spectrogram of a factory noise.

Horn sound was also considered, sound with an original sampling frequency of 44.1 KHz. Using sptool, the time domain information shows a maximum peak of 1.000e+00 at 0.089s,a minimum peak of -1.000e+00 at 0.102s, a peak to peak of 2.000e+00,mean of -6.303e-05,median of -3.052e-05 and RMS value of 2.404e-01.It's Power Density increases from -70dB(0Hz) upward till -40dB (513.4Hz) and there was a decrement downward till-92dB and further till at-94.5dB(min. PD) both at 22050Hz.It's spectrogram shows partial higher energy distribution at 4KHz below and with partial low energy distribution at 5KHz to 20KHz above.

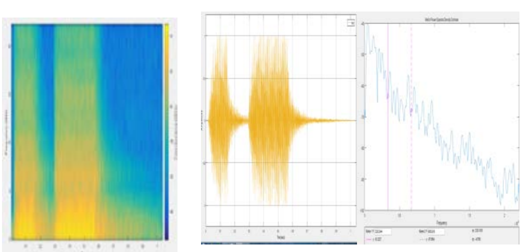


Fig.4.3: (a) A spectrogram of a horn sound (b) The PSD display of a horn sound(c) Time domain waveform of a horn sound.

Generator noise otherwise known as white noise was acquired by a high-quality analog noise generator. It was considered as one of the uncontrollable environmental sound with an original sampling frequency of 44.1 KHz. Using sptool, the time domain information shows a maximum peak of 7.708e+03 at 2.491s,a minimum peak of -7.954e+03 at 2.354m, a peak to peak of 1.566e+04,mean of 1.001e+02,median of 1.010e+02 and RMS value of 1.509e+03.Its PD increase from 22.2dB to 25dB at 0Hz and decrease downwards to 23dB(137Hz). The PD virtually remains constant from 137Hz till 8694Hz and eventually decreases to 21dB at 10 KHz and decreases downwards to minimum PD 18dB at 10 KHz. Its spectrogram shows a uniform higher power intensity across the entire plot.

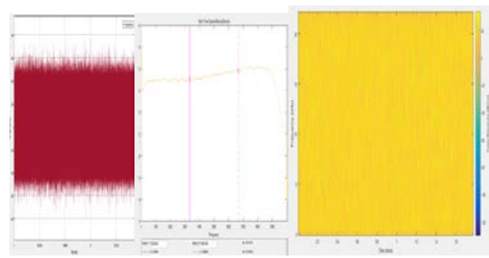


Fig. 4.4: (a) Time domain waveform of a generator noise (b) The PSD display of a generator noise(c) A spectrogram of a generator noise.

V. DISCUSSION AND CONCLUSION

A priori knowledge of the characterization of a noise or environmental signal, can be use to subtract or cancel the noise from a noisy signal in order to enhance the quality and intelligibility of a speech signal [8].

In signal characterization, asking which domain is more important is equivalent to asking the question which is more important in life food, air or water, it all depends on the approach intend to use in developing the algorithm for speech enhancement algorithm. It is very difficult or impossible to develop an algorithm that works efficiently in all types of noise environment as reported in [9]. In order to narrow the scope of acoustic signals considered for characterization due to the uncountable source of environmental sounds that interferes with speech signal during communication and in an attempt to develop a generic noise reduction algorithm which can perform efficiently in all types of noise environments, uncontrollable environmental sounds with high level of noise intensity was considered. In order to develop an ideal speech enhancement algorithm that works efficiently in all types of noise environments, it was proposed in [9], an autonomous system has to be develop, that works both online and offline. The offline mode occurs, when the user is not making a call, it was proposed the system should be able to recognize the current environment where the user his and based on this environment, when online (making call), a speech enhancement algorithm automatically enhance the speech based on the present environment. The technique proposed in [9] will be further investigated, in order to realize, an ideal speech enhancement algorithm that works very well in all noise environments.

References

[1] *P. Mathys , " Sound in Time and Frequency", Telecommunications 1.[Online] Available:* https://ecee.colorado.edu/~mathys/ecen1200/sound/sounds 2006_2pp.pd.*[Accessed: June 06, 2017].*

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

[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] S. Chu,"Unstructured Audio Classification for Environment Recognition", Proc. of the 23rd AAAI Conf. on AI.,L.A.,USA,2008, pp.1845-1846. [Online] Available: *https://www.aaai.org/Papers/AAAI/2008/AAAI08-312.pdf*.[Accessed: June 16, 2017].

[5] National Instruments, "Categorizing Signals (Advanced Signal Processing Toolkit)" [Online] Available:http://zone.ni.com/reference/en-XX/help/372656C-01/lvasptconcepts/categorizing_signals/ [Accessed: June. 28, 2017].

[6] P.C, Louizou ," Introduction", Speech Enhancement Theory and Practice, 2nd Ed., New York, CRC press 1999, pp.1-6.

[7] H. C. Mahendru and R. Khanna, "Spectral Analysis of Various Noise Signals Affecting Mobile Speech Communication" IJSER., India, vol. 5, April-2014, pp 1087-1094.

[8] C.N. Prabhavathi and K. M. Ravikumar", Acoustic Noise Characterization and Classification Using Statistical Properties:[Online] Available: *https://www.ijetae.com/files/Volume4Issue6/IJETAE_0614_99.pdf.* [Accessed: June. 28, 2017].

[9] S. Selva et. al.," A review on speech enhancement algorithms and why to combine with environment classification" Int. Journ. of Modern Physics C. vol. 25, No. 10 2014, pp 1430002-1430017. [Online] Available: http://www.worldscientific.com/doi/abs/10.1142/S0129183114300024[Accessed: July. 28, 2017].

[10] J. Salamon, C. Jacoby and J. P. Bello, "A Dataset and Taxonomy for Urban Sound Research", *22nd ACM Int. Confr. on Multimedia, Orlando USA, Nov. 2014.*

[11] D. Ellis, "Noise", Sound Examples. [Online] Available: https://www.ee.columbia.edu/~dpwe/sounds/noise/ [Accessed: June. 28, 2017]

[12] D. H. Johnson ," Signal Processing Information Base (SPIB)".[Online] Available: http://spib.linse.ufsc.br/noise.html

[13] www.convertfiles.com

[14] National Instruments, "Aliasing and Sampling at Frequencies above the Nyquist". [Online] Available: http://www.ni.com/white-paper/3000/en/ [Accessed: July 20, 2017].

[15] Mathworks,"sptool", [Online] Available: https://www.mathworks.com/help/signal/ref/sptool.html [Accessed: July. 23, 2017].

[16] Emba, "Lab 3. Spectral Analysis in Matlab - Part II", EE289 Lab Jan., 2012.[Online] Available http://www.emba.uvm.edu/~gmirchan/classes/EE270/EE275/2012/Matlab/Lab3 [Accessed: July 23, 2017].

[17] Kay, Steven M.," *Modern Spectral Estimation: Theory & Application".* Prentice Hall Signal Processing Series, 1987.

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

Olawore Jelili Oladayo his currently a lecturer at the Department of Computer Engineering, University of Maiduguri, Borno State, Nigeria. He had is first degree in Computer Engineering from Ladoke Akintola University of Technology in 2004 and a master degree in Electronic

ISSN 2229-5518

Communication and Computer Engineering from the University of Nottingham, 2012. He is a member of the Nigeria Society of Engineer (NSE), a register engineer with Council for Regulation of Engineering in Nigeria (COREN).His area of research includes energy efficiency, high speed communication system, soliton transmission system, machine learning and embedded systems. He is currently enrolled for his PhD at Ladoke Akintola University of Technology in Nigeria.

Dr. Christopher Umerah Ngene received is M.Sc. degree in computer engineering from Kiev Institute of Civil Aviation engineers, Kiev, Ukraine in 1992 and Ph.D degree in computer systems and components from the Kharkov National University of Radioelectronics, Kharkov, Ukraine in 2012. His dissertation was on embedded verification infrastructure for an assertion-based HDL-code for digital systems on chip. Models and methods were developed for assertion analysis using transaction graph models and tabular methods. A hardware accelerator was developed using VHDL. He is currently an Associate Professor in the Department of Computer Engineering, University of Maiduguri, Nigeria. His research interests include and not limited to Computer architecture, Design and test of Computer Systems/Systems on Chips (SoCs), Design for testability and Computer Networks.

Dr. Abdulfattah Adelani Aboaba holds a PhD degree from the Department of Electrical and Computer Engineering, International Islamic University,Malaysia.His research interest are in image surgery(IGS)/Computer Aided Surgery(CAS),embedded systems, image segmentation and analysis, computation oncology and high performance computing(HPC).His work experience spans both academic and industry. He is currently a senior lecturer in the Department of Computer Engineering, University of Maiduguri, Nigeria and member of the Faculty Board of Postgraduates. He is equally a visiting senior lecturer at the Federal University of Technology, Minna. Abdul Fattah is a member of the Nigeria Society of Engineer(NSE),a register engineer with Council for Regulation of Engineering in Nigeria(COREN), a member of Institute of Electrical and Electronic Engineer(IEEE), a member of Engineering in Medicine and Biological Society(EMBS), a member of American Association of Science and Technology(AASCIT) and a senior member of Asian-Pacific Chemical, Biological and Environmental Engineering Society(APCBEES).He has over twenty-five publication in both national and international journal and conference proceedings and served as an editor for 2015 IEEE International Conference on Electronics, Communication and Networks (CECNet 2014).One of its publication was recently rated among the most viewed articles in the Germany-based online research evaluation platform, Research gate.