ANALYSIS OF MMSE SPEECH ESTIMATION IMPACT IN WEST SUMATRA'S NOISES

Suardinata, Vetho Sayuthi and Zainul Efendy

STMIK Indonesia Padang, Padang, Indonesia

ABSTRACT

This study aimed to estimate the original voice signal which is interrupted by noise with MMSE method based on Wiener filter. The Wiener filter is classified as the MMSE estimator studied by previous researchers. The study assessed the voice signal input count down by European woman that are distorted by two types of noises, the one is noise based on the outdoor location, sound of siren firefighter, and the other is the indoor location noise, which represented by noise in lecturer room in campus. The two process signal is estimated by MMSE estimator which approximated by Wiener filter that must have founded and counted the covariance of each signal processes are related to the system. Thus the researchers tried to estimate both types of sounds noisy European woman are due to interference noise source of fire fighter and faculty room and assess its impact in the form of graphs vote against a function of time, the pattern shape of the signal and SNR.

KEYWORDS

Voice Signal, MMSE method, Wiener filter, covariance

1. INTRODUCTION

In general conversation, the impact of disturbances caused by noise. This noise impact consists of various kinds of noise sources, and depending on the location. In past research have been widely observed or assessed Europe woman's voices signal [18], but for the study of noise interference on the geographical aspects in the region of western Sumatra have not been studied, it is encouraging to do research on these aspects, by dividing the two categories, namely: sound noise as a result of indoor noise sources on STMIK INDONESIA PADANG campus, which consists of voice conversations as well as some fan noise from the effects of wind, and noises which can not be defined. While the other noise category of other locations (outdoor), a swab sirens firefighter, due to the main street in one of the Padang city highway in West Sumatra frequently traversed by one of the main source of noise, the sound of a siren firefighter. Woman's voices are taken from the primary sample in the internet, and have the voices of Europe woman speaking because in previous studies also examined the objects of speaking Europe in [18], with the sentence pronounced to give a message about the countdown from number 10 to zero. The second type of noise that is considered to represent noise that occurs in West Sumatra on mention above is added to the mathematical sum in order to give effect to the disruption in the European woman's voices. Thus the researchers tried to estimate both types of European woman noisy sounds, due to noise source interference of fire fighter and faculty room and assess its impact in form of amplitude graphic to function of time, pattern or shape of the signal and Signal to Noise Ratio (SNR).

2. PROBLEM BACKGROUND

2.1 Defintion of Sound

Sound is a form of energy that is emitted by the vibrating section and reaches the ear and causes the sensation of hearing through the nerve tissue. [1]. It generated by all vibrating parts but not everything can be heard. Frequency limit of hearing is from 20 Hz to 20,000 Hz. Noise intensity is measured in decibels (Db) [2, 3]. Voice sound has the distinction of sound kind itself, that consist of phonetic factor of physiology pronunciation and comprehension as well as acoustic and phonology factor of pattern combination of the voice sound consist of phonotactic regularity series and aspects of the impact of conversation context [11].

2.2 Improving The Quality of Sound

The important method of improving sound quality by eliminating background noise and suppress echo. Technique improves sound quality may varies, [5]. In general, the noise can be added, multiplied and convoluted, noise can be classified in a narrow frequency band, the broad frequency band, stationary or non-stationary. [12].

Most of the processing for improved sound quality belonging to the summation method noise, wide-banded and stationary noises. The ability to improve the sound quality by trying to get a clear signal of the original sound of the noisy voice signal is important in order to get a comprehensive understanding of the voice. Sound improvement algorithm can be connected with voice recognition or voice coding at the end of the process, so as to increase the speed of recognition reliability [5]. And the impact on voice coding can reduce the number of bits that are important in signal coding. Examples of useful applications in the system to improve the quality of a sensor in this case the microphone is a mobile telephone systems are used in environmental noises. Methods in improving the sound can be either Weiner filter method, Kalman filter, MMSE, Model restoration autoregressive model coefficients, the hidden Markov model and search models pitch. While the system is in the sound quality improvement methods are as follows: segmentation and windowing in the speech signal, Discrete Fourier Transform, spectral analysis, correlation inter-frame and intra-frame, and a voice model. For the method in many channels can be Acoustic Echo Remover System engineering, Multiple Input Multiple Output (MIMO), Beam Forming Microphone Array, Adaptive Noise Removal.

2.3. Definition of Noise

Noise is unwanted sound, [1]. The noise can also be defined as any unwanted noise and instead of the usual natural phenomena. [6]. Noise comes from the Latin "nausea" which means unwanted sound or a loud noise, do not like, or do not expect, [20]. Noise is the disruptive noise caused by vibration of the material. Noise has two sources, natural sources such as wind, lightning, volcanic eruptions, waves lapping the beach, bird, animal noises, the movement of the wind, the movement of ocean currents, or waterfall.

Artificial sources, sound machines, plants, jams, household appliances, and transportation systems. The transport system is a major source of noise pollution in urban areas. Construction of buildings, Construction of buildings, highways, and public roads, causing a lot of noise, the use of air compressors, bulldozers, oader, trucks, and breaking the asphalt road, [8]. Noise is different from pollution else, noise usually has sources those more than one in any location, so it is not easy to control, unlike the environmental pollution such as pollution of water, air and soil, pollution originating from the plant can be processed, reduced, or the plant was closed, while the

noise found in homes, schools, roads, offices, and other places where humans can not control it, [4]

Noise of industry also added the condition that has been unwelcome from noise pollution. Loudspeakers, plumbing, heating boilers, generators, air-conditioners, fans, and air cleaners, adding the presence of pollution noise, noise, road congestion, air, noise from rail, noise, construction, noise in the industry, noise from consumer products, means loudspeakers public announcements and instructions, and firefighters[4]

2.4 MMSE Estimation

On Error! Reference source not found., explained that if X and Y are random variable with joint distribution, defined $E\{.\}$ stated expectations, with distribution of a joint across random variables involved in the calculation, $E = \{X|Y\}$ stated the average estimate conditional X if Y values are known. So that the conditional variance of Y, be

$$var\{X|Y\} = E\{(X - E\{X|Y\})^2 |Y\} \tag{1}$$

The average estimation is optimal condition on the average value of the square, So that the MMSE score to estimate the X given Y to shape the conditional variance of the mean, is,

$$MMSE(X|Y) = E\{var\{X|Y\}\} \tag{2}$$

MMSE score may be considered as a function of Signal to Noise Ratio SNR for each statistical distribution of X, P_X , another meaning of the MMSE score is as functional distribution of the input signal X, P_X. When the input of the value of the Gaussian with average m and variance σ_x^2, mathematically expressed by

$$X \sim N(m, \sigma_x^2) \tag{3}$$

$$MMSE(X, SNR) = \frac{\sigma_x^2}{1+\sigma_x^2} \tag{4}$$

2.5 Estimating the Noise

In order to obtain the estimated value of original sound clear signal obtained from noisy sound signal, can be estimated by looking at the source of its noise, if the noise is stationary noise detection can be performed explicitly (SE) as in [10] or the method further as minimum statistical methods (SM) by [16]. For noise is not stationary, there are six methods that can be done, that is the way that has been classified codebook (CK) by [19], the method is suitable for certain types of noise and too complex to be applied to mobile phones, hearing aids, and the tool is limited by the low level of complexity. Approach TFD-subspace (ST) on [14], and the method of reduction in the level of complexity of the algorithm ST with high-resolution-ST (SR) on [15]. In the ST and SR, the sound signal is modeled by less rank methods.

Another estimation is obtained by weighted noise estimation (EDB) to implement the RSD estimating noise with noise on the heuristic weighting function [13]. The function of the optimal weight can also be obtained by measuring the average error minimum squares or minimum mean - squared error (MMSE) the [18], the MMSE method (ME) also has to consider aspects of bias based compensation noisy signal model whereas the compensation aspect bias heuristic way obtained on the method of [17].

Biased estimator obtained by the following formula below:

$$B = \frac{\sigma_W^2}{E_Y\{E\{N^2 \mid y; \xi\}\}} \tag{5}$$

σ_W^2 = RSD ideal noise, E_Y = expectation of Y, Y = coefficient TFD noisy voice, N = the magnitude of the noise in the number of polar, W = coefficient TFD noise, y = a random variable noisy sound signal, ξ = a priori SNR. Error log symmetrical distribution in the study examined the tracing of noise aspect, using symmetric calculation error log that can be searched by the following formula, the error log symmetric,

$$\log - eror = \frac{1}{IK} \sum_{k=1}^{K} \sum_{i=1}^{I} \left| 10 \log_{10} \left[\frac{\sigma_W^2(k,i)}{\hat{\sigma}_W^2(k,i)} \right] \right| dB \tag{6}$$

I = amount of beacon frames

σ_W^2 obtained by performing smoothing on periodograms as a function of time and valuable mathematically exponential window.

$$\sigma_W^2(k,i) = 0.9\sigma_W^2(k, i-1) + 0.1|w(k,i)| \tag{7}$$

2.5.1 Signal to Noise Ratio SNR

Performance improvement of sound quality can be tested using PESQ and segmental SNR

$$SNR_{seg} = \frac{1}{I} \sum_{i=0}^{I-1} T \left\{ 10 \log_{10} \frac{\|x_t(i)\|^2}{\|x_t(i) - \hat{x}_t(i)\|^2} \right\} \tag{8}$$

where $x_t(i)$ and $\hat{x}_t(i)$ said frame-time I x_t clear voice signal and the voice signal is improved or enhanced x_t

2.6 Identity Sound

In [7], speech recognizer automatically by computer speech recognition is the process by which voice signals are automatically converted into a row of words in a text relating to or in accordance with the sound of the words in question. In the development of voice recognition-based Hidden Markov Model has achieved high performance in an environment that has been conditioned. Application of voice recognition used in various types of environments. This recognition performance will be reduced if the environmental conditions are used as a reference and the environmental conditions where the application of voice recognition is different.

The purpose of the automatic voice recognition is to develop techniques and systems that enable the computer to accept voice input. Voice recognition problem frequently encountered is the time to translate or convert voice signals into text form by computer. In generally computer users want their conversation that is in the form of voice signals can be translated or converted into text form by computer.

2.6.1 Extraction Mechanism of Sound

On the mechanism of voting is a process to obtain useful information from the signal by ignoring, eliminating, or without the effect of the repetition or the presence of unwanted information.

In practice, while eliminating unwanted information may also lose some of the resources that are useful in the process. In an option may contain changes the signal into a form suitable model used according to the classification system is appropriate. In the construction of the automated selection, the following is desirable factors, namely:

1. Good or high distinction in between each word sub-class.

2. Can work on various types of speakers

3. The decline in the quality of voice signal is invariance due to disruption of the communication channel or as a result of noise.

Objective processes that occur in the automatic speech recognition is to find a clear and strong nature of the correlation acoustically, which parameters can be computed or estimated associated with the processing of the signal waveform.

Another thing that is dividing the sound signal into these parameters. It aims to generate an appropriate interpretation of the sound signal. The processes involved in selection is already included the process of converting the signal into digital form for example in signal conditioning for further processing, measurement of some important properties of signals such as energy and frequency response, the development of the measurement process into several measurement signals important in attempt to translate the message of the voice signal and statistical requirements that form the signal vector of the observation signals.

The ultimate purpose of the selection was to divide the sound signal into the various components that can be identified acoustically or to obtain a decrease in the speed of a set of changes functions in order to get the amount of computing that allows to process the voice signal.

3. PROPOSED METHOD

In this study, the recorded voices files in the form of * .wav, then added by a Gaussian noise likely such as outdoor firefighter noise and indoor lecturer room noise and, then processed by wiener filter of MMSE method estimator, After that, the noisy voice signal which has already estimated with noise eliminated by MMSE estimator, then further analyzed the results - the results of processing these signals. At the end of the study was concluded on the performance of the research methods used in this study as a whole. The steps performed in the testing of this noisy voice signal can be seen as follows:

The sound of the original signal is taken from the voice signal as shown in this list:

```
m=wavread('emergency012a.wav');
p=m.';
t=[0:1:541855];
p2=wavread('codo.wav');%//p2 sinyal
p2=p2.';
p1=[zeros(1,420524) p];%//p1 sirine
p3=p1+p2;
plot(t,p3); %// data sinyal p2 dgn p1 sebagai sirine
wavwrite(p3,41000,16,'codoDerauSiren.wav');%// data terderau noise

%// estimasi sinyal codo + sirine

cx11=cov(p2);cn11=cov(p1);
x11=(cx11/(cx11+cn11))*p3;
wavwrite(x11,41000,16,'codoDerauSirenEst.wav');
%//SNR dan BER
vp2=var(p2);
vp1=var(p1);
snr11=vp2/vp1
```

```
%//
hitung=0;
for i=1:541856
if (p1(i)==p2(i))
hitung=hitung+1;
endif
endfor
hitung

%//DERAU RUANGAN DOSEN STMIK PRTAMA KALI SEJAK 1985
mr=wavread('ruangdosenstmik.wav');
mr=mr(1:1:2145024,1);
pr=mr.';
t=[0:1:2145023];
%//t=[0:1:541855];
p2=wavread('codo.wav');%//p2 sinyal
p2=p2.';
p2=[zeros(1,1603168) p2];%//p1 sirine
p4=pr+p2;
plot(t,p4); %// data sinyal p2 dgn p1 sebagai sirine
wavwrite(p4,41000,16,'codoDerauSRD.wav');%// data terderau noise

%//
cx12=cov(p2);cn12=cov(pr);
x12=(cx12/(cx12+cn12))*p4;
wavwrite(x12,41000,16,'codoDerauSRDEst.wav');
```

```
%//SNR dan BER
vp2=var(p2);
vpr=var(pr);
snr12=vp2/vpr
%//
hitung2=0;
for i= 1603168:2145024
if (pr(i)==p2(i))
 hitung2=hitung2+1;
endif
endfor
hitung2
```

4. SIMULATION RESULT AND DISCUSSION

The result of our simulation is discussed here, the simulation consist of the voice signal input that count down from ten to zero by European woman. This European woman voice signal is added by the outdoor location noise and compare with indoor location noise, in this research the noise is lecturer room.

After the adding process in speech programming in computer software that imitated as a real condition that happened in outdoor and indoor condition. The result of adding final processing is obtained by computer software programming.

The overall signals are explained by this bellow discussion.

4.a. Voice Signal Input Count Down by European Woman

This European woman voice signal is captured by computer software programming as shown in figure 1.
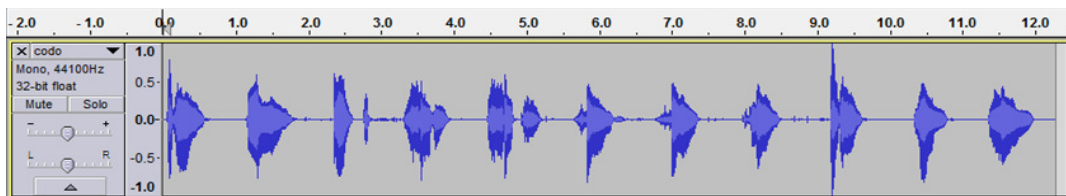

Figure 1 European woman signal in time series.

4.b. Voice Signal that are Distorted by Noises

The voice signal are discussed are distorted about by two types of noise impact,

1. Noise based on the outdoor location

2. Noise based on the indoor location

The example of outdoor location noise in Padang city in the west Sumatra noises is represented by Padang city roadway noise, is sound of siren firefighter signal. The firefighter siren signal in time series is obtained from internet as shown figure 2
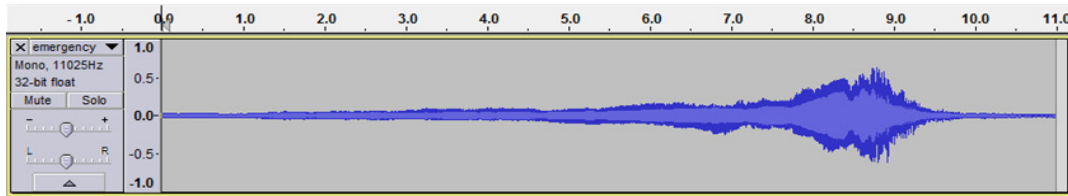
Figure 2 Firefighter Siren signal in time series.

The siren noisy voice signal is shown as figure 3. It is noises firefighter signal with European woman voice signal in time series, the noise itself is outdoor type noise which is represented as one of kind of West Sumatra noise.
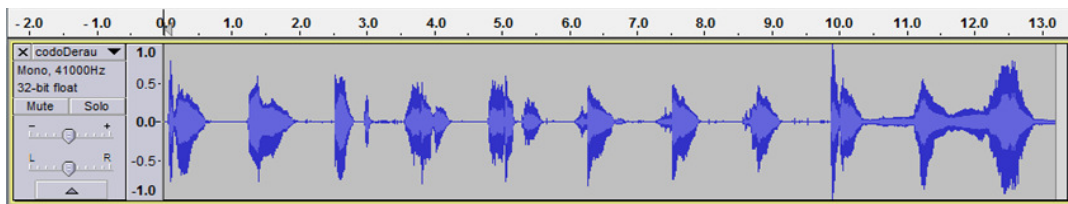

Figure 3 Noises firefighter signal with European woman voice signal in time series

4.c. Indoor Location Noise,

The other noise is indoor location noise, which represented by noise that occurs in lecturer room in STMIK INDONESIA PADANG campus that located in West Sumatera. As shown in figure 4.
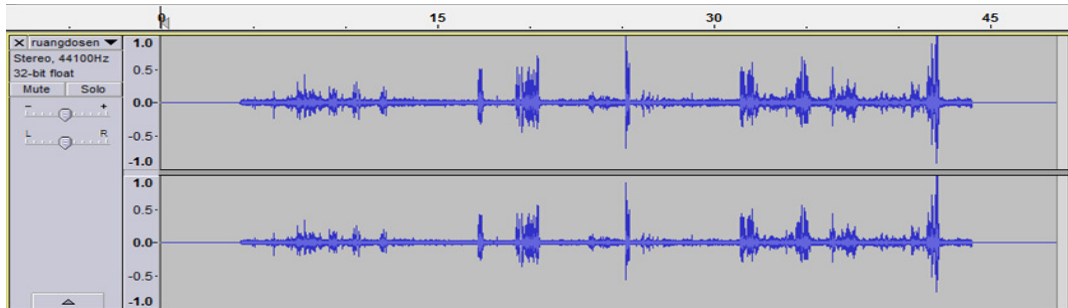

Figure 4 Lecturer room noise in time series

The indoor lecturer room noisy voice signal with European woman voices signal in time series that obtained by adding process of the woman voices signal and the indoor room noise is computed by computer software programming as in figure 5.
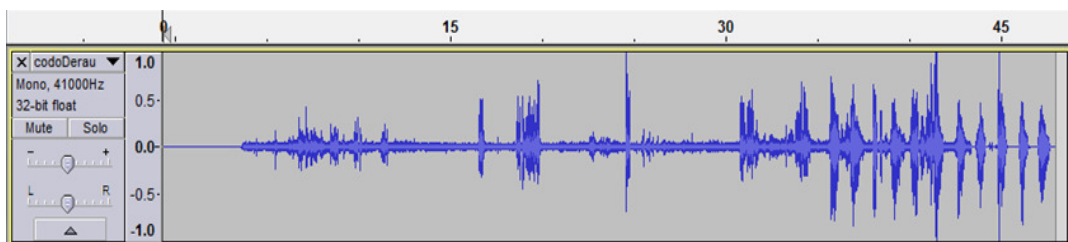

Figure 5 Noisy indoor lecturer room signal with European woman voices signal in time series

18

4.d The Estimated Signal

The two process signal is estimated by MMSE estimator which approximated by Wiener filter,
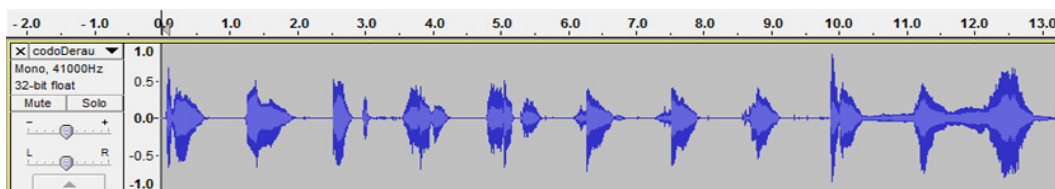
a. Estimated siren noisy voice signal

Figure 6 Estimated noisy sirine signal with European woman voices signal in time series

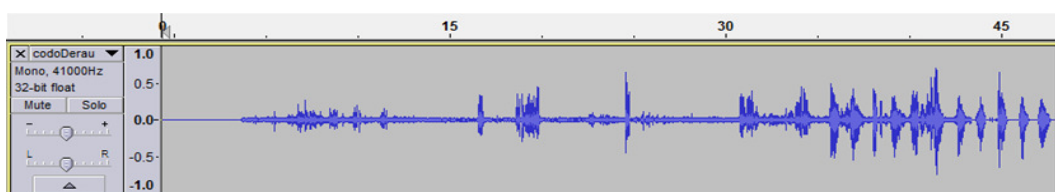### b. Estimated indoor lecturer room noisy voice signal

Figure 7. Estimated noisy indoor lecturer room signal with European woman voices signal in time series

The presented table of experimental systems is,

Table 1. Estimated noisy sirine firefighter and indoor lecturer room with European woman voices signal in time series experimental systems

Parameter	Values
Original speech signal 1	541856 samples
Siren noise signal	541856 samples
Original speech signal 1 covariance	+ 0.011696
Siren noise signal covariance	+ 0.0019665
Original speech signal 2	2145024 samples
Indoor lecturer room noise signal	2145024 samples
Original speech signal 2 covariance	+ 0.0029545
Indoor lecturer room noise signal covariance	+ 0.0015825
SNR 1 (Original speech signal 1 to Siren noise signal)	5.9477
SNR 2 (Original speech signal 2 to Indoor lecturer room noise signal)	1.8670
SNR 1 (dB)	7.743491 dB
SNR 2 (dB)	2.711443 dB
Signal System	Mono-Stereo, 11025 Hz, 41000 Hz, 44100 Hz, 32 bit float,
Equal Symbol 1	8,872 to 2,145,024
Equal Symbol 2	4,529 to 541,856

5. CONCLUSION

The original sound is then adding by the indoor source noise sound and outdoor source noise sound given, also had an impact on the quality of the sound. The Estimated MMSE process is already done versus the noise sound. From the research results MMSE estimation was able to reduce noise impact to the original sound European woman voice signal as described by MMSE estimation result shown in the form of the wave in the simulation, specially in figure 6 and 7 above.

The original speech sample has value as long as 541,856 samples that adding with siren noise signal at the same of the length of sample, the covariance of speech signal is 0.0116 in outdoor and 0.0029 in indoor, the siren noise covariance is 0.0019, and the indoor noise covariance are 0.0015. The SNR value variated because of the relative distance length from the source to the measurement point. The outdoor SNR that has been noted reached 7.74 dB, and indoor SNR is 2.71 dB.

REFERENCES

[1] Al-Fifi, H.I., Al-Azri, A.F. and Al-Aetby, T.K. 2006. "Noise Pollution", King University Press, Saudia, Riyadh".

[2] Al-Hassen, Sh.I. 2011. "Environmental Pollution in Basra City", Ph.D. thesis, College of Arts, University of Basra".

[3] Al-Hassen, Sh.I. 2013. "An Assessment of Noise Pollution and Associated Health Impacts at Selected Schools in Basra City, Southern Iraq". Journal of Basrah Research.

[4] Al-Kelaby, A.S. 2013. "Indoor and Outdoor Air, Water & Noise Pollution in Sammawa City, Iraq". Ph.D Thesis, College of Arts, University of Basra, Iraq.

[5] Alugonda Rajani, and Soundarya .S.V.S, August 2016. "A Review on Various Speech Enhancement Techniques" , International Journal of Advanced Research in Computer and Communication Engineering, IJARCCE, ISSN (Online) 2278-1021, ISSN (Print) 2319 5940, ISO 3297:2007 Certified Vol. 5, Issue 8.

[6] Bronzaft, A. 2004. "Noise Pollution", in (Editor Richard M. Stapleton), Pollution A to Z, Vol. 1, Macmillan Reference, New York.

[7] Bhupinder Singh dan Dr. Joginder Singh, January 2016, "A Study of Feature Extraction for Automatic Speech Recognition", Volume 6, Issue 1, ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering Research Paper. IJARCSSE.

[8] Duha Saleh Karem, Hamzah Abdulhasan Kadhim, Hamid T Al-Saad, 2016, "Assessment of Noise Pollution in West Qurna-2 Oil Field Southern Iraq", Journal of Scientific and Engineering Research.

[9] Dongning Guo, Yihong Wu, Shlomo Shamai (Shitz), Sergio Verdú, APRIL 2011, "Estimation in Gaussian Noise: Properties of the Minimum Mean-Square Error", IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 57, NO. 4, 2371.

[10] J. Sohn, N. S. Kim, and W. Sung, January 1999 "A statistical model-based voice activity detection," IEEE Signal Processing Lett., vol. 6, no. 1, pp. 1–3.

[11] John Nerbonne, 2001, "Speech Sounds Introduction to Linguistics for Computational Linguists".

[12] Kotta Manohar and Preeti Rao, 2005, "Speech Enhancement In Nonstationary Noise Environments Using Noise Properties".

[13] M. Kato, A. Sugiyama, and M. Serizawa, 2001, "Noise Suppression With High Speech Quality Based On Weighted Noise Estimation And MMSE STSA".

[14] R. C. Hendriks, J. Jensen, And R. Heusdens, March 2008 "Noise Tracking Using DFT Domain Subspace Decompositions".

[15] R. C. Hendriks, R. Heusdens, J. Jensen, And U. Kjems, 2009, "Fast Noise PSD Estimation with Low Complexity," In IEEE Int. Conf. Acoust., Speech, Signal Processing".

[16] R. Martin, July 2001, "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics," IEEE Trans. Speech Audio Processing, vol. 9, no. 5, pp. 504–512.

[17] R. Yu, 2009, "A Low-Complexity Noise Estimation Algorithm Based On Smoothing Of Noise Power Estimation And Estimation Bias Correction," In IEEE Int. Conf. Acoust., Speech, Signal Processing.

[18] Richard C. Hendriks And Richard Heusdens Jesper Jensen, 2010, "MMSE Based Noise PSD Tracking With Low Complexity", Delft University Of Technology, The Netherlands.

[19] S. Srinivasan, 2005, "Knowledge-Based Speech Enhancement", Ph.D. Thesis, KTH.

[20] Singh, N. and Davar, S. 2004. "Noise Pollution-Sources, Effects and Control. Journal of Human Ecology. Subspace Decompositions," IEEE Trans.Audio Speech And Language Processing, Vol. 16.

AUTHORS

Suardinata, he is received the Diploma III 1999 in Information Management at AMIK Riau, Indonesia, Bachelor Degree in Information Engineering from STMIK Riau, Indonesia, and Master Degree in Information Technology from Universitas Putra Indonesia, Padang, Indonesia. Currently he is a Ph.D student in the Dept. of Computer System and Communication, Faculty of Computer Science and Information System, Universiti Teknologi Malaysia, Johor Bahru Malaysia. He has been working as Lecturer at STMIK Indonesia Padang from 2005 in the Dept. of Computer Science and Information Systems, STMIK Indonesia Padang. His research interests include Multimedia and Voice over IP network, Signal and Image Processing, Network Security, Traffic Engineering and Quality of Service issues in IP networks, Wireless Ad-Hoc Networks, and Distributed Systems.

Vetho Sayuthi, he is received the Bachelor 2005 in Telecommunication Engineering at STT Telkom Bandung, Indonesia, and Master Degree in the same college at 2008. Currently he has been working as Lecturer at STMIK Indonesia Padang from 2016 in the Dept. of Computer Science and Information Systems, STMIK Indonesia Padang. His research interests include Signal, Speech, Image and Multimedia Processing, and Network Engineering.

Zainul Efendy, he is received the Bachelor 2004 in STMIK Indonesia Padang, Master Degree in the UPI YPTK Padang in 2012. currently he is lecturer in STMIK Indonesia padang at information system department . His research interests include database management system