# A Study on Improving the Overloaded Speech Waveform to Distinguish Alcohol Intoxication using Spectral Compensation

Won-Hee Lee [#1], Seong-Geon Bae [*1] and Myung-Jin Bae [**3]

[#1] Information and Telecommunication of Dapartment, Soongsil University
369, Sangdo-ro Dongjak-gu, Seoul

[1] vbluelovev@ssu.ac.kr

[2] sgbae123@empal.com

[3] mjbae@ssu.ac.kr

***Abstract*—In this study, speech characteristics from before and after alcohol intoxication has been comparatively analyzed through speech analysis to obtain the degree of intoxication along with its parameters. However, the overload of the input signal must be considered prior to anything else. When distinguishing intoxication under overload conditions, an error cannot be avoided. Therefore, distinguishing the level of intoxication must be conducted by using a method that is not significantly influenced by overload. Thus, reliability of distinguishing the degree of intoxication was enhanced by pre-processing the overload to compensate the spectrum. The SNR and intoxication distinguishment rate resulted comparatively better at about 2 and 5% overload clipping rate, however, resulted in an error and increase at 10% overload clipping rate.**

**Keyword- Waveform coding, non-uniform Sampling, Peak and Valley, Adaptive filter, Zero-crossing**

## I. Introduction

Drinking is an unavoidable part socializing in life. Relationships can be restored or destroyed when drinking socially. Moreover, alcohol can be consumed for uniting an organization, but excessive drinking can also ruin it. Hence, alcohol can either be a gift or a curse. Until recently, breathalyzers have only been used for traffic management. However, nowadays, breathalyzers are installed in smart phone applications, navigation systems, and on the steering wheels so cars cannot be started when being under influence. Such measuring instruments only work within close proximity. On the other hand, measuring levels of intoxication through speech makes long distance quantifications possible. Long distance measurements through speech is especially favorable for railways, ships, aircrafts, cars, and other methods of transportation where close proximity is not achievable.

At such long distances, an overload of input signal must be considered. Overload refers to a value that exceeds the load of a machine or equipment where a distortion occurs at the signal processing circuit. An overload usually occurs when the input microphone is too close, input level is too high, or when the voice is too loud after being intoxicated. An error will occur when distinguishing the level of alcohol intoxication under such overload conditions. Thus, an intoxication distinguishment method that is not affected by overload must be used.

In this paper, spectral compensation was employed under overloaded signals to distinguish levels of intoxications from overloaded speech waveform. Chapter 2 will deal with characteristics of overload signals and propose a method for compensation, chapter 3 will examine the experiment and results. Chapter 4 will discuss the results and future research directions.

## II. Characteristics of Overloaded Signal

Overload is commonly known as a value that exceeds the range of what a device can measure. Speech signal is overloaded when the input microphone is too close, background noise is too loud, or input level is set high. Additionally, overload occurs depending on the input amplitude of the microphone and when the environment is irregular under wireless communication environments. Various reasons exist for overload, however, because an overloaded signal is a value that exceeds the normal range, its waveform has a clipping phenomenon at the top and bottom peaks (Fig. 3-1). Characteristics of overloaded signal from a pure tone will be analyzed before analyzing the overloaded speech signal. Figure 1 shows an overloaded 350Hz pure tone within time domain.

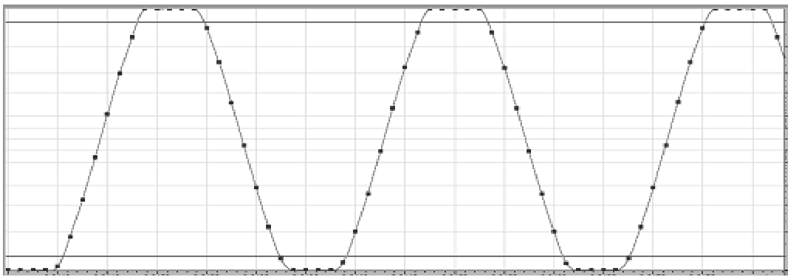
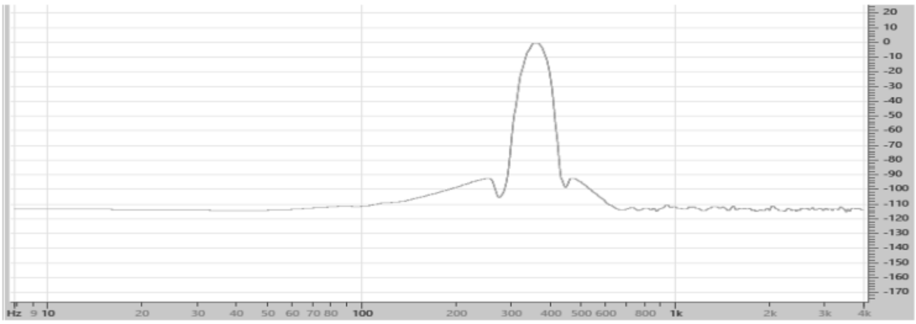


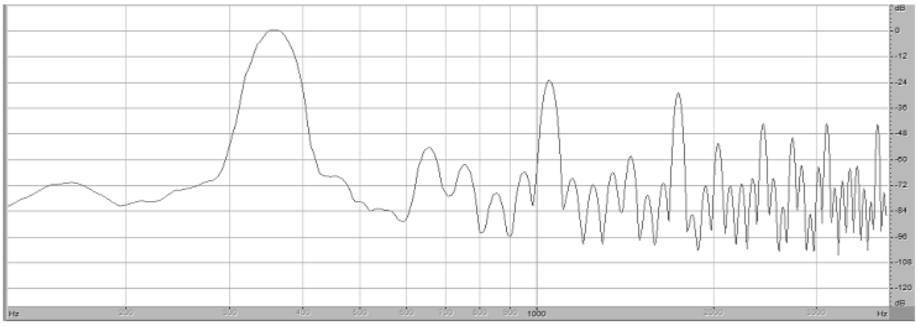
Figure 1. Time domain waveform of 350Hz pure tone overloaded by 20%

As shown in figure 1, there is a limit to expressing all the signals levels from the incoming input levels when an overload occurs. Therefore, as the amplitude exceeds the maximum level, a clipping phenomenon occurs at the top and bottom. Hence, having a similar form to that of a square pulse.

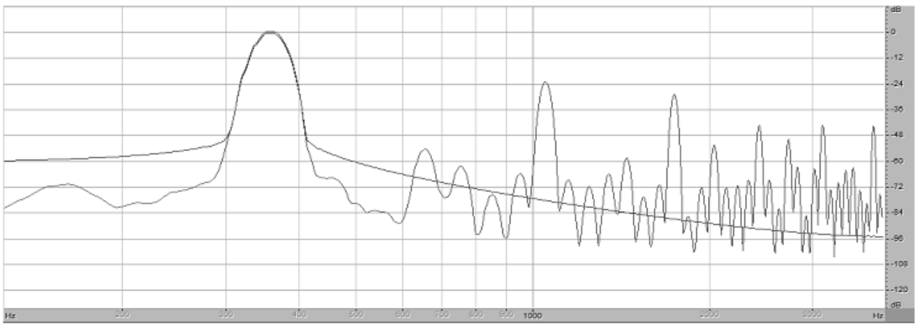
Figure 2 shows spectral changes when overloading a pure tone of 350Hz. Figure 2-a shows the spectrum of 350Hz pure tone, whereas, figure 2-b shows the spectrum of a 20% overloaded pure tone. Lastly, figure 2-c shows a comparison of the two figures above.




(a) Spectrum of 350Hz pure tone



(b) Spectrum of 20% overloaded 350Hz pure tone



(c) Comparison of 350Hz pure tone and 20% overloaded pure tone

Figure 2 Spectral changes from 350Hz pure tone to overload

Figure 2-b shows harmonic distortions in the event of overloading the pure tone. Harmonic distortion occurs due to the change of waveform, like that of the square pulse, caused by clipping the top and bottom peaks (Fig.1). Figure 3 shows the spectrogram of the pure tone and overloaded pure tone.

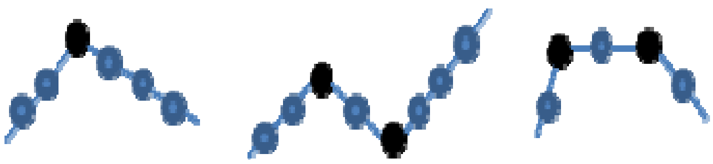

(a) Spectrogram of 350Hz pure tone


(b) Spectrogram of 20% overloaded 350Hz pure tone
Figure 3 Spectrogram change when overloading 350Hz pure tone

As shown in figure 3-b, harmonic distortion can be easily observed compared to that of figure 3-a. Moreverover, it is confirmed that the harmonic overtone is generated only at a numerical system (1,3,5,7...). Therefore, a low-pass filter is needed to mitigate harmonic distortions during restorations.

## III. OVERLOADED SPEECH SIGNAL COMPENSATION AND PROPOSED METHOD

In this paper, speech signal can be classified into unvoiced and voiced. Sound pressure energy of unvoiced is loud but consists of irregularities, however, the sound pressure energy of voiced is loud and quasi-periodic. Therefore, voiced and unvoiced must be restored differently due to their differences. Unvoiced has a high zero-crossing rate resulting in static noise and peaks and valleys; the sound pressure energy is high but is not used in restoration because it is less affected by overload. On the other hand, clipping signals are produced when voiced is overloaded. [4] Hence, it is significantly influenced by overload. Figure 4 shows a common peak and valley as well as clipping by overload.



(a) peak (b) peak and valley (c) Clipping by Overloaded
Figure 4 typical peak and valley and clipping by overload

Figure 1 (c) represents overload signal due to high sound pressure during input. Such signal is an excessive distortion of the low frequency, and doubles the average frequency scale in terms of frequency domains. Method of improving the overload signal is processed into two phases; low and high frequencies. First, the low frequency component is configured in the vowels. High frequency components are what causes the peaks and valleys especially affecting voiced to be overloaded. Figure 3-5 shows the schematic of improving overload signals to distinguish intoxication. First, the overloaded input signal is normalized to 50%. Normalization is conducted due to synthesizing the low and high frequencies after improving them separately. Next, about 500Hz of low pass filter is used to improve the low frequency. As shown in figure 4-4, harmonic distortion occurs by clipping when overloaded. Furthermore, in order to improve a clipped signal, the low pass filter is needed to closely resemble the shape of the square pulse for smoothing to create the peaks and valleys. The next step consisted of a cut-off frequency between 1000Hz and 3500Hz and used the band pass filter. Such is necessary to improve the high frequency within the signal. However, overload causes harmonic distortions which in turn increases the high frequency energy at above 3500Hz, thus is the reason why band pass filter is used. Next the two signals are synthesized. Lastly, about 50% of the signal that passed through the band pass filter is utilized and synthesized to improve the high frequency component.

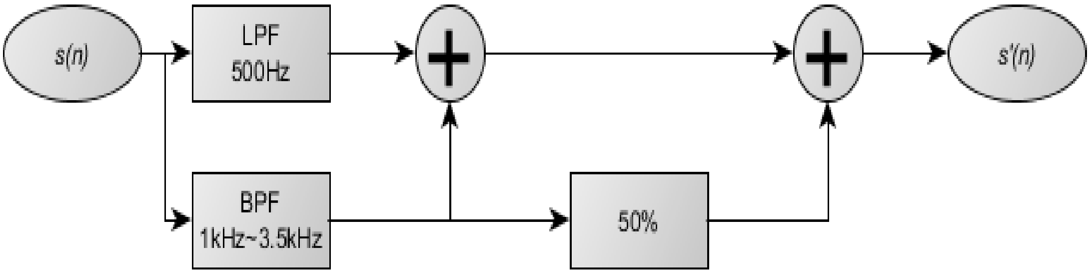
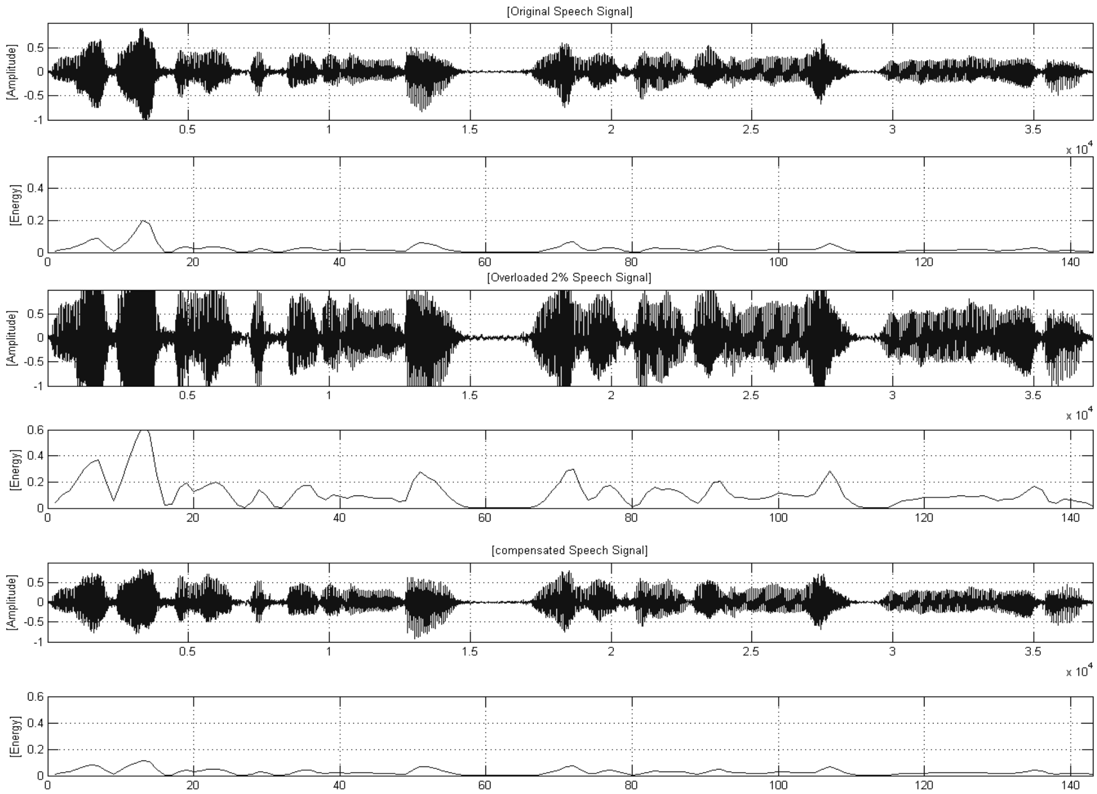


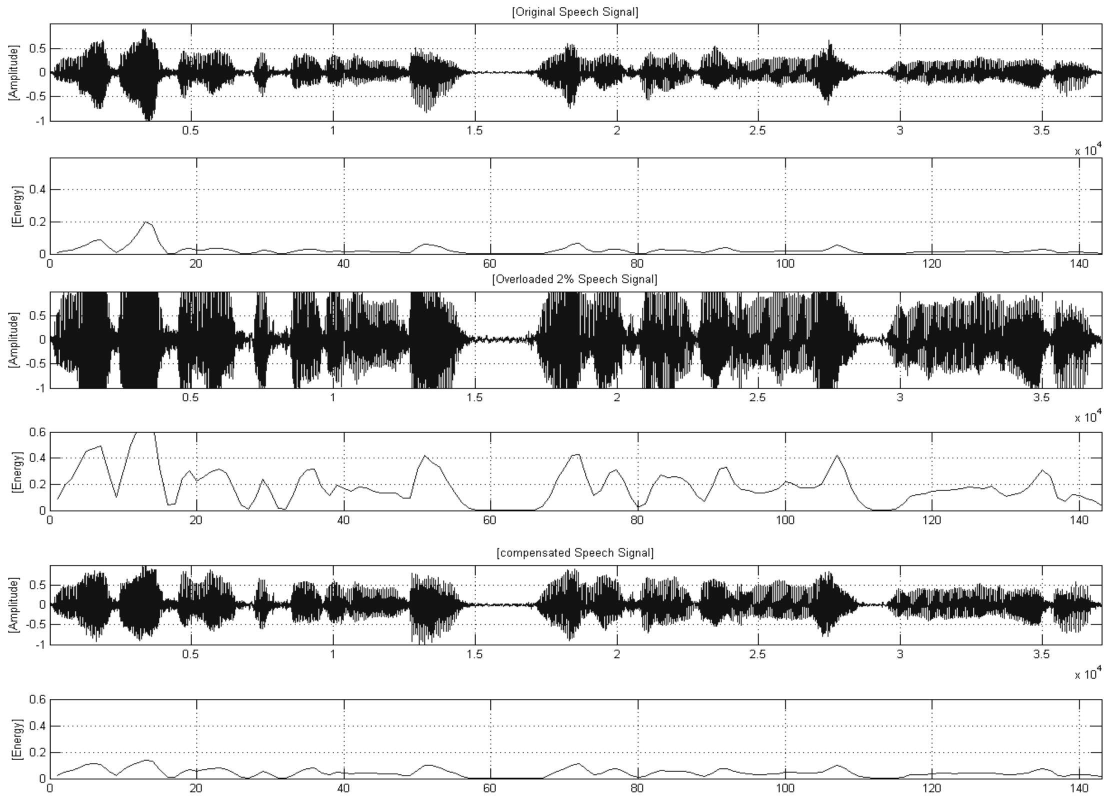

Figure 5 Schematic diagram of the proposed algorithm

## IV. RESULTS AND DISCUSSION

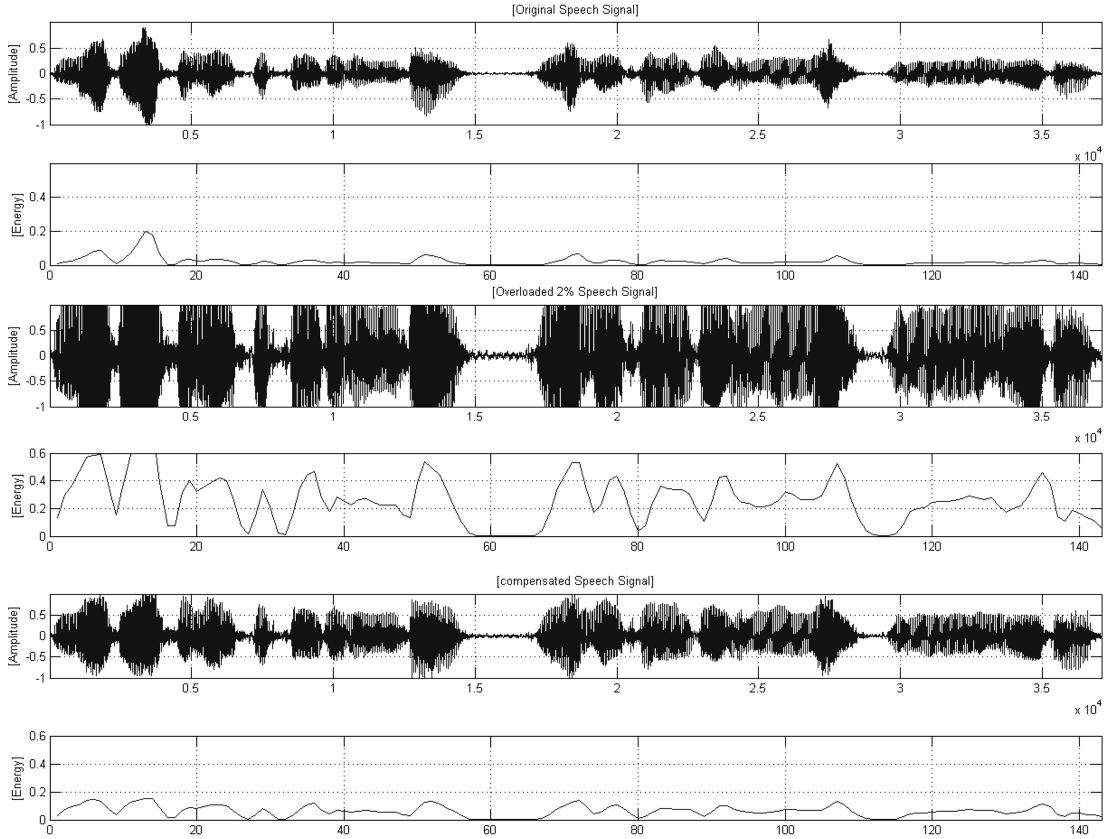
Figure 6 shows a normal input signal and clipping rate in respect to speech waveform and compensated speech waveform.




(a) energy envelope curve of compensated speech signal at 2% clipping rate

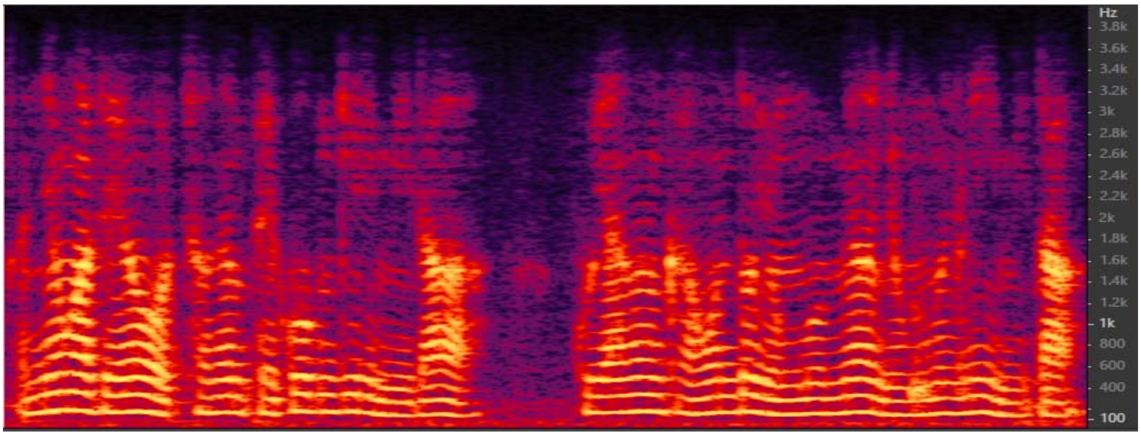



(b) energy envelope curve of compensated speech signal at 5% clipping rate

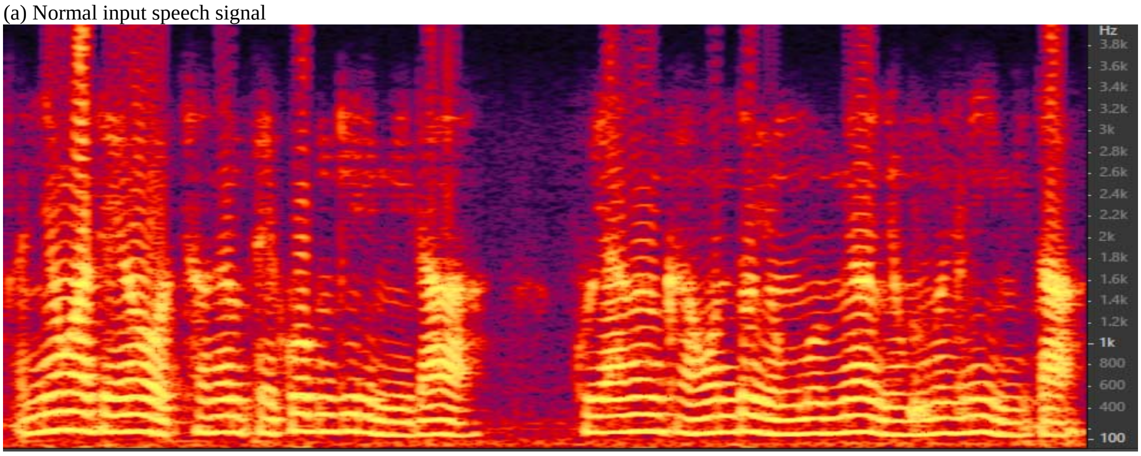



(c) energy envelope curve of compensated speech signal at 10% clipping rate

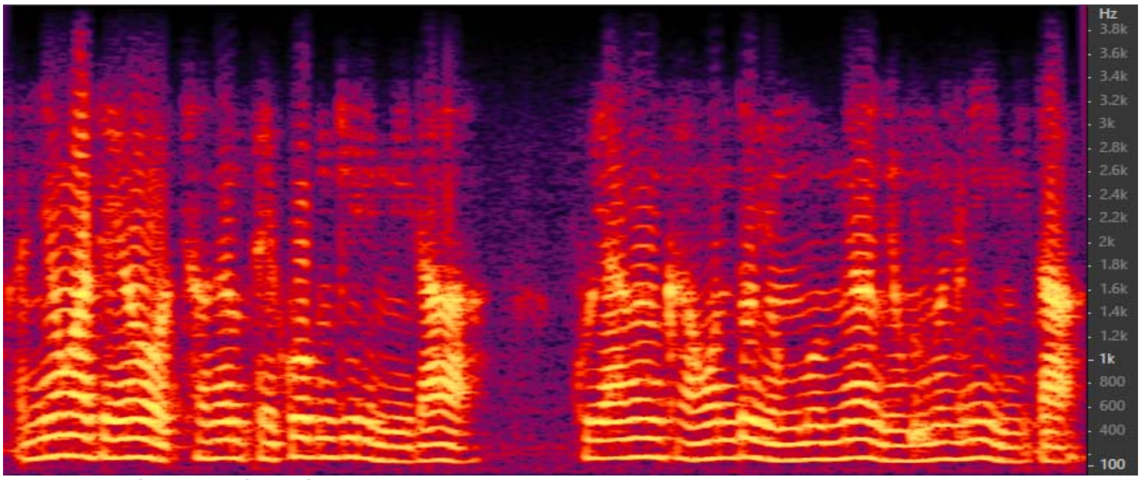
Figure 6. Energy comparison of compensated overload signals




(a) Normal input speech signal




(b) overload input speech signal




(c) compensated input speech signal

Figure 7 Spectrum comparison of compensated overload signal

The energy envelope curve shown in figure 6 is very similar to each other. However, when observing each frame of the energy envelope curve of compensated overload signal, an error is produced before 20 frames. It can be conjectured that the clipping rate is increased due to overload and could not be compensated completely. Figure 7 shows spectrum comparisons of restored overloaded signals.

Comparatively analyzing SNR in respect to the raw signal and each clipping rate, 2% clipping compensation showed that the SNR value of the raw signal did not differ whereas 5% showed a difference of about 0.8dB. However, compensating 10% clipping waveform showed an average of 1.9dB decrease compared to the raw signal.

Table 1 shows levels of alcohol intoxication through improving overloaded speech signals after drinking followed by the use of previous algorithm. Previous algorithm compares the First Formant and Fourth Formant energies using LPC. The waveform closely followed the intoxication probability of the raw signal until about 5% clipping, however, at over 10% clipping, the probability could not be relied upon.

Table 1 Comparison of clipping compensation before and after alcohol intoxication and standard intoxication judgment rate (LPC tilt comparison)

| Speech Sample | Raw signal intoxication probability [unit : %] | 2% clipping compensation | | 5% clipping compensation | | 10% clipping compensation | |
|---|---|---|---|---|---|---|---|
| | | Before | After | Before | After | Before | After |
| Sample 1 | 63 | 48 | 64 | 63 | 73 | 75 | 71 |
| Sample 2 | 55 | 70 | 66 | 64 | 55 | 67 | 70 |
| Sample 3 | 63 | 58 | 64 | 65 | 62 | 63 | 58 |
| Sample 4 | 50 | 58 | 66 | 66 | 52 | 67 | 57 |
| . . | . . | . . | . . | . . | . . | . . | . . |
| Sample 19 | 75 | 65 | 79 | 75 | 74 | 75 | 63 |
| Sample 20 | 52 | 76 | 59 | 55 | 53 | 46 | 54 |
| Sample 21 | 62 | 66 | 59 | 60 | 61 | 59 | 61 |
| Sample22 | 67 | 66 | 61 | 55 | 62 | 54 | 64 |

## V. CONCLUSION

Close range breathalyzers only work within certain proximities raising problems regarding human rights and accidents. Therefore, long range method of distinguishing levels of alcohol intoxication is needed. One of many ways of distinguishing alcohol intoxication from far distances is through speech. However, despite being under influence, it is difficult to check alcohol intoxication when the person denies this fact [4]. Furthermore, distinguishing the level of alcohol intoxication is even more difficult if the microphone is too close and there is an overload signal.

In the occurrence of a signal overload, the input levels are clipped due to the speech signal exceeding the maximum input value of the device. Such clipping results in harmonic distortion. This paper thus performed experiments to improve overloaded speech waveform to distinguish intoxication through spectral compensation. In order to compensate signal overload, both low and high frequencies must be considered separately to avoid harmonic distortion. Hence, the overloaded input signal must be normalized to about 50% because low and high frequency components are improved separately and then synthesized. To improve low frequency components, a low-pass filter of about 500Hz must be passed. This is followed by a cut-off frequency from 1000Hz to 35000Hz, succeeded by the use of band-pass filter; the two signals are then synthesized. Lastly, about 50% of the signal that has passed through the band-pass filter is synthesized for high frequency compensation. Through the use of such technique, distinguishing intoxication levels at about 2 and 5% showed promising results whereas, errors occurred at over 10%. Therefore, future research will compensate the overloaded signal algorithm and improve noise within communication environments along with implementation to high input levels.

## REFERENCES


[1] S.G. Bae, M.J. Bae, “A New Speech Coding using Harmonics Emphasis Filter,” ISAAC 2013, AACL Vol. 1, pp. 43-44, November 2013.

[2] Changjoon Jung, Seonggeon Bae, Myungsook Kim and Myungjin Bae, “Speech Sobriety Test Based on Formant Energy Distribution,” IJMUE Vol.8, No.6, pp.209-216, 2013.

[3] S.G. Bae, M.S. Kim, and M.J. Bae, “A New Hybrid Non-Uniform Coding with Low Bit Rates for Sound Signal in Near Field Communication,” 2013, IEEE ICTC, pp. 1039-1040, ICTC 2013.

[4] Seonggeon Bae, Myungsook Kim, and Myungjin Bae, “On Enhancement Signal Using Non-uniform Sampling in Clipped Signals for LTE Smart Phones ”, IEEE ICCE-berlin(2013), pp.125 - 126.

[5] S.G. Bae, M.S. Kim, and M.J. Bae, “On Enhancement Signal Using Non-uniform Sampling in Clipped Signals for LTE Smart Phones,” 2013, IEEE ICCE-berlin, pp.125-126, ICCE-berlin 2013.

[6] S.G. Bae, M.J. Bae, “On A New Hybrid Speech Encoder using Variable LPF,” ASTL Vol. 21, pp.275-278, April 2013.

[7] S.G. Bae, H.W. Park and M.J. Bae, “On a New Enhancement of speech Signal using Non-uniform Sampling and Post Filter,” ICHIT2012, LNCS 7425, pp.723-729, August 2012.

[8] Marcieli Bellé, Sílvia do Amaral Sartori, Angela Garcia Rossi. “Alcoholism: effects on the cochleo-vestibular apparatus”. Rev Bras Otorrinolaringol. 73(1):116-22. 2007.

[9] M.J. Bae, D.S. Kim, H.Y. Jeon and S.G. Ann, "On a new predictor for the waveform coding of speech signal by using the dual autocorrelation and the sigma-delta technique," IEEE Proc. of ISCAS'94, vol 6, No. 3, pp.261-264, June 1994.
[10] Behne, D. M., & Rivera, S. M. (1990). Effects of alcohol on speech: Acoustic analysis of spondees. Research on Speech Perception, Progress Report, 16 (pp. 263–291). Bloomington, IN: Speech Research Laboratory Department of Psychology, Indiana University.

## AUTHOR PROFILE

Won-Hee Lee He received the M.S. degree in Telecommunication & Information from Soongsil University in 2014. He is currently the under Ph.D. degree at Soongsil University.

Seong-Geon Bae He received the Ph.D. degree in Telecommunication & Information from Soongsil University in 2014. He is currently the Professor of the Dept. of Broadcasting Sound & Visual at Daelim University.

Myung-Jin Bae He received the Ph.D. degree in Electronic Engineering from Seoul National University in 1987. He is currently the Professor of the Dept. of Information & Telecommunication at Soongsil University. He has authored and coauthored more than 300 journal articles and conference papers. His research interests are Speech Signal Processing, Audio Processing and Speech Communications