A Judgment of Intoxication using Hybrid Analysis with Pitch Contour Compare (HAPCC) in Speech Signal Processing

Seong-Geon Bae*1
1Professor, School of Software Application, Kangnam University, Gyunggido, Korea.

Won-Hee Lee2
Research Scholar, Telecommunication & Information Department, Soongsil University, Seoul, Korea.

Myung-Jin Bae3*
Professor, Telecommunication & Information Department, Soongsil University, Seoul, Korea.

(*Correspondence author)

Abstract
Research on the estimation of body condition through voice analysis has been carried out to date. One of the reasons why your body suddenly changes is drinking. Drinking causes changes in the body and changes in the voice. Characteristics such as pitch, formant, and change of sound size are studied as parameters to discriminate between alcohol and alcohol, and voice before and after drinking is discriminated through high frequency and low frequency change of voice signal. In this paper, we investigated the alcohol discrimination method by comparing the change of voice signal before and after drinking using the changing vocal cord characteristics. By applying the pitch container to the musculature, the results of the drinking survey obtained various characteristics.

Keywords: Intoxication, pitch, formant, cepstrum, frequency

INTRODUCTION
There are various kinds of drinking in the world. Looking at the renowned alcohols consumed or manufactured, one can understand life or culture of certain country or region. Alcohol is something that we cannot ignore in our life and is consumed in various places by different methods.

Alcohol makes people become excited and plays a positive role of making people become closer to each other. However, it decreases one’s thinking ability, resulting in reduced self-control. As a result, there are many accidents such as sexual assault, murder, or drunk driving take place. Drunk driving is one or the major accidents that happen after the drinking and may kill someone like murder. As numbers of drunken drivers are increasing, government is trying to enforce crackdown or related regulations. However, due to the issues such as lack of manpower or violation to human rights, there are limits to this. Currently, crackdown on drunken drivers on road is relatively easier but it becomes extremely hard when it comes to railroad, air, or sea. To enable drunk driving check in different places, this shall not only be done in short-distance but also in long-distance areas as well. One of the easies method to identify changes in the body is to analyse the voice. Voice is made from inside to outside the body and is delivered to a long distance. If we can find out changes in voice before and after the drinking, it will allow highly effective crackdown on drunken drivers[1][2][3].

This research tries to find voice before and after drinking by applying variation in the pitch. Pitch is the basic frequency in voice signal and is very sensitive to the body changes. As there are many changes in the body when one drinks alcohol, visible changes are made after the drinking compared to the before. If we can emphasize these changes in the pitch, we will be able to check intoxication more accurately. In Chapter 2, it discusses about the general method of processing the voice signal and followed by this, it describes special techniques of this research in Chapter 3. In Chapter 4, it looks into the test and result, lastly, makes the final conclusion in Chapter 5.

CHARACTERISTICS OF OVERLOADED SIGNAL
In speech signal processing, pitch is an essential element in speech processing systems, such as formants. In the speech synthesis, the pitch representing the characteristics of the excitation source determines the naturalness of the sound, and the speech characteristic is used for speaker identification because it has the characteristic characteristics of the speaker during speech analysis. These features have been used for a long time in speech synthesis, analysis and recognition is. Therefore, various algorithms for pitch extraction have been proposed. Depending on the method used, it can be broadly divided into time domain, frequency domain and hybrid domain analysis.

Figure 1. A general pitch detection in each domain
Autocorrelation and AMDF techniques are used in the time domain. In the frequency domain, FFT, spectral flattening, SAMDF method, etc. are widely used. In the hybrid area, the Cepstrum method is widely used. Figure 1 is a graph showing the pitch extraction method for each region. Since the pitch extraction method in the time domain is directly processed in the time domain, no conversion process is required to another domain, and the pitch is extracted by emphasizing the periodicity of the speech waveform. Although the processing speed is faster than other areas, when the signal is passed through the characteristic transmission channel, or when background noise is added, the error is increased because it is extracted by sum, difference and comparison logic. In the frequency domain, the amplitude spectrum is obtained by frequency-converting the speech signal, and the pitch is extracted by emphasizing harmonics. Since the frequency conversion is performed, it is possible to extract a pitch more robust against noise and more accurate than the time domain. However, the process of converting to frequency is time consuming and complicated. Today, the development of DSP chips has reduced the computational time burden in the frequency domain. In the hybrid domain, the pitch is extracted considering only the advantages of time and frequency domain. It can save calculation time in time domain and overcome problems such as background noise and transition period in frequency domain. In this paper, Cepstrum is applied to compare pitch changes before and after drinking.\[1][2][4][11]\n
PROPOSED METHOD
Drinking causes a variety of physical changes. The voice changes as well. First, before the alcohol goes into the stomach, it is first contacted with the mouth and esophagus. Drinking alcohol with high alcohol content for a long time or a strong stimulus to the oral mucous membranes or esophagus wrapped around the esophagus and cause dehydration. In addition, drinking alcohol after the diuretic effect of alcohol can cause dehydration of the mucous membranes of the larynx, which acts on the vocal cords to change the voice. Second, the flexibility and elasticity of the vocal fold mucosal tissue is reduced, and when the hypoglossal pressure is applied, the complete vocal cord opening does not occur, resulting in the air leaking when talking after drinking, which reduces the clarity of pronunciation. Third, the increase in the viscosity of the vocal cord muscles and the incompleteness of the glenohumeral strength increase the minimum pressure required for vocalization. As a result, it is necessary to start and maintain the vocalization, which leads to an increase in the vital capacity of the vocalizers, do. The result is not only a change in speech energy but also an early termination of the sentence.\[3][4][5][6][7][8]\n
\[ E_{VF} = \frac{1}{M_v} \sum_{n=-\infty}^{\infty} \sum_{m=1}^{M_v} s^2(n + m) \]  \hspace{1cm} (1) \n
Where \( M_v \) is number of valid frames.

Finally, after drinking, the ability to control the vocal organ is reduced, resulting in a tongue twisting sound when pronouncing, and the pronunciation is often crushed, resulting in less accuracy. In addition, the phenomenon of nasalization is also prominent [6]. These phenomena mean that the characteristic parameters of the speech signal have changed. In this paper, we compare before and after drinking using characteristics of vocal fold muscle changes for clarity of pronunciation in drinking. Figure 2 is a block diagram of the proposed method. Hybrid analysis using both time domain and frequency domain parameters of drinking judgment was used to solve the problem of analysis by each domain.

**Figure 2.** A block diagram of proposed method
Figure 3. An entire of waveform before drinking

Figure 4. An effective of waveform before drinking

Figure 5. A contour of PCCS per frame before drinking
Figure 5. A contour of PCCS per frame after drinking

Figure 6. An entire of waveform after drinking

Figure 7. An effective of waveform after drinking
When I do not drink alcohol, I have the characteristic of pronouncing every syllable in order to pronounce correctly when speaking. Conversely, drinking alcohol causes dehydration due to alcohol rather than before drinking, which makes it difficult to vocalize before drinking. Because it is necessary to give strength to the muscles before drinking, it is characteristic that the pronunciation is cut off accurately after drinking, rather than continuing to speak. To determine the change of vocalization before and after drinking, peak point is obtained by using capstrum, and the average change rate of size value is compared. Figures 3, 4, and 5 show the changes in the original speech waveforms, the effective speech waveforms, and the peak points per frame. Figures 6, 7, and 8 show the changes of original sound waveforms, effective speech waveforms, and peak points per frame after drinking. Comparing Figures 5 and 8, the average rate of change of peak size was 0.0022 before drinking and 0.0009 after drinking. The overall envelope of the picture shows that even after drinking, the change is similar to before drinking, but the instantaneous rate of change per frame is greater than after drinking. When the average change rate is large in alcoholic beverage, it is possible to change the sound without difficulty when pronouncing it, so that it can be pronounced correctly, but after drinking, the change of the sound does not occur so much that it can not be pronounced incorrectly.

CONCLUSION
Drinking is an essential part of people's lives. Proper drinking will help not only people but also social life. On the other hand, if you drink a lot of alcohol, the person's control ability will be lowered and you will have difficulty in correct action. In this paper, we analyzed the voice before and after drinking by using the change of the peak amplitude of the capstrum of the voice signal. It is possible to pronounce the pronunciation correctly without difficulty before drinking. On the other hand, after drinking alcohol causes dehydration of the vocal muscles, preventing proper pronunciation. Using these characteristics, the change of the sound before and after drinking was extracted and compared. It is expected that this study will be useful for various body changes in the future. The characteristics of the body according to drinking were different according to age and sex, but focused on detecting general characteristics. Our experiments are based on the detection of common features and the development of algorithms for alcohol discrimination through relative analysis.

REFERENCES

AUTHOR PROFILE
Seong-Geon Bae He is currently the Professor of the School of Software Application at Kangnam University.
Won-Hee Lee. He is currently the under Ph.D. degree in Telecommunication & Information from Soongsil University.
Myung-Jin Bae. He is currently the Professor of the Dept. of Information & Telecommunication at Soongsil University.