International Journal of Multimedia and Ubiquitous Engineering
Vol.8, No.5 (2013), pp.395-402
http://dx.doi.org/10.14257/ijmue.2013.8.5.39

A Study on Voice Sobriety Test Algorithm in a Time-Frequency Domain

Geumran Baek* and Myungjin Bae

Information and Telecommunication Department, Soongsil University, 369 Sangdo-ro DongJak-Gu, Seoul, 156-743, Republic of Korea

judybaek@hanmail.net, mjbae@ssu.ac.kr

Abstract

Voice is an important tool to identify individuals. In recent years, the voice can determine personality, age, and health status as well as physical conditions such as height and stature. The purpose of this paper is to develop an algorithm to determine whether when one is intoxicated by using the voice; characteristic parameters of the intoxicated voice signal were extracted by comparing the voices before and after drinking using several speech analysis methods. First, the change was observed in the body, glottis and vocal tract after drinking in order to examine the influence of drinking on phonation. In addition, differences of voice signals of a variety of the speakers were examined before and after drinking in order to find characteristic parameters for intoxicated voice signal that enable a judgment about whether or not one is drunk. The result of this study is expected to prevent drinking and driving, if applied to the development of the intoxication judgment system.

Keywords: *Voice, intoxicated voice signal, the development of the intoxication judgment system*

1. Introduction

Most accidents caused by drunk driving are serious that they even claim the lives of the drivers. Nowadays, accidents caused by drinking from the marine and aviation flight are frequent in addition to car accidents; the government are making several efforts to prevent the drink and driving. The current ways of drinking crackdown is to directly measure the degree of the driver or navigator, among which ways there are the way by the breath meter and blood sampling assay method [1]. Transport, railways, ships, aircraft, *etc.*, however, are often difficult to stop while operating for measurement in close proximity. Therefore, alcohol measurement through speech analysis can be an alternative to the two methods mentioned earlier, in order to determine the flight drinking.

The simplest way to measure the remote ship operator for intoxication is to predict whether the navigator is drunk through communication. It may be difficult to determine intoxication right away if the original voice is mistaken as drunken voice if judged in a very subjective way, or if the navigator denies being intoxicated despite of having drunk. Therefore, this study is aimed to develop an indirect and objective algorithm of speech analysis for drivers in order to judge the degree of intoxication by analysis of the voice of the operators during the communication of the flight.

---

* Corresponding Author : Myungjin Bae, Professor of Information and Telecommunication Department, Soongsil University, mjbae@ssu.ac.kr.

ISSN: 1975-0080 IJMUE

Chapter 2 examines the change in voice after drinking; Chapter 3 described the intoxication decision algorithm using the extraction of parameters characterized by drinking, which is proposed in the present study. In addition, Chapter 4 describes the experiments and results and Chapter 5 concludes the paper.

2. Changes in Voice after Intoxication

When we drink, the alcohol contacts the oral cavity and esophagus first before being absorbed into the gastrointestinal tract. If we drink undiluted strong liquor like soju, whiskey or kaoliang chiew for an extended time, they cause strong stimulus to mucous membrane that surrounds the esophagus or the top layer (epidermis) and dehydration. Further, the diuretic effect of alcohol after drinking can cause the dehydration of the larynx mucosa, which phenomenon can act to the vocal cord to change the voice. It also reduces the flexibility and elasticity of the vocal cord mucosa tissue, thus may cause incomplete opening/closure of the vocal cord when subglottal pressure is applied, and degrades the clarity of pronunciation due to the air leak when talking after drinking. The increase in the viscosity of the vocal cord muscle and the imperfection in the opening and closing of the glottis exalt the phonation threshold pressure, which causes the increased lung capacity and hard breathing because the voice user needs power to initiate and maintain the phonation [2, 3]. As a result, the change in the energy of voice occurs and premature termination closure of the sentence is caused during pronunciation.

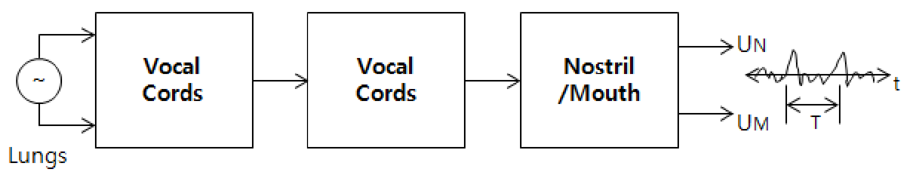

Figure 1. Voice Generation Model

Figure 1 shows the block diagram of the lower glottis system consisting of the lungs, bronchial tract and respiratory organs, in other words, the organ system generating the energy creating the voice [5]. Voice is an acoustic waveform emitted from the system when the air is released from the lungs and vibrated by the point of stricture in the vocal tract. The change of phonation after drinking means the displacement of the pitch and formant, the characteristic parameters of the voice signal described. The pitch frequency is the parameter that contains the characteristics of the tune and rarely changes depending on the changes of the body, while formant frequency is the parameter that contains the characteristics of the tone and changes relatively easily due to the changes of the body and emotion, so it is estimated to be an important way to determine intoxication by the speech signal [4, 6]. However, pitch is also an important parameter of the voice, so it is worth being investigated if it plays any role as a parameter of the drinking voice.

Several noticeable and important parameters can be found from the two voice signals by comparing the voice before and after drinking, carefully considering the above-mentioned characteristics. For example, the change in voice energy, change of voice pitch, degraded accuracy of pronunciation, change in speed and duration of pronunciation, early closure of the end of the pronounced sentence, and the change in the pitch and formant are the characteristics from which parameters can be extracted to determine the status and degree of drinking.

3. Characteristic Parameter of Intoxicated Speech

The next section will describe the extraction of the characteristic parameters for intoxicated voice in both time-domain and frequency-domain, which is proposed in the present paper.

3.1 Extraction of time-domain parameters

Several characteristics could be obtained from the drinking voice signals observed in the time-domain. First, the deviation of the loudness appeared worse. This is somewhat coherent to the common fact that the voice of an intoxicated person does not maintain its appropriate volume but gets louder or smaller.

Second, the sound volume does not tends to be maintained uniform in smaller sound phonation after drinking rather than in bigger sound phonation. In fact, accurate control of the vocal organ becomes more difficult when its movement is smaller, for the movement of the vocal organ, including the brain becomes slower after drinking. For the small phonation, the width of the movement of the lungs and the vocal cord should inevitably be reduced, so it is more difficult to maintain such narrow energy constant from the sluggish brain and the vocal organ [8].

Finally, regarding the speech waveform after drinking, the energy distribution ratio is increased for the higher level compared to the lower-level energy distribution ratio. This is considered to be the phenomenon caused by degraded auditory function after drinking, which makes the speaker to say louder, opening the mouth wider, thus increasing the lung capacity.

Figure 2 shows the flowchart for time-domain process to predict the extent of drinking. As explained earlier, the attention was paid on the change of the energy of the voice signal, taking account of the spirometric change, deterioration of hearing ability and the changes of speech sounds, caused by the physical changes after drinking. First of all, the energy(En) was obtained for each frame, followed by the comparison of the energy(Ed1,Ed2) of the front and rear part of the speech signal. In addition, significant intervals were obtained, followed by the energy for each interval. Finally, the energy ratio(ER) was obtained between the intervals and was used as the drinking voice parameter.

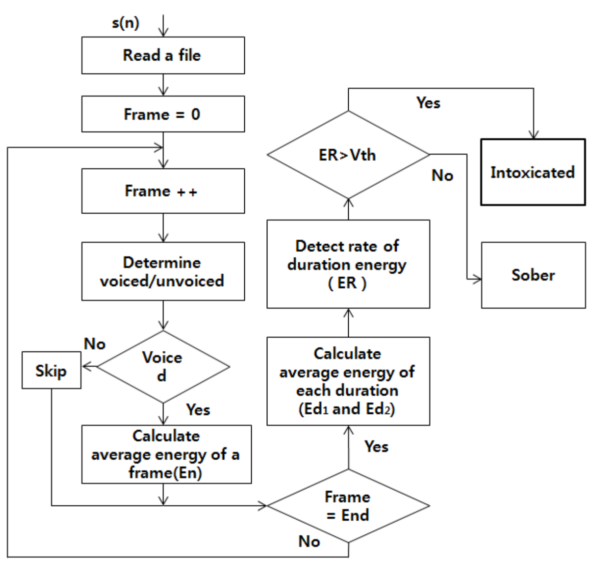

Figure 2. Flowchart of time-domain prediction process for intoxication

3.2 Extraction of frequency-domain parameters

The feature of the intoxicated drinking in the frequency domain is the nasalization phenomenon of voice caused by drinking alcohol. In addition, the voice of the intoxicated speaker gets louder in spite of oneself due to the degraded auditory function. This means that the drinker phonates with the mouth wide open, which feature is the change in the lowest formant [8].

In addition, drinking makes the change in the position of the tongue, thus affects the second formant. In other words, the anterior tongue position make the second formant bigger, the posterior tongue, smaller. Therefore, the voice characteristics after drinking can be defined as the relationship of the slope of the first and fourth formant (F14) and the slope of the second and fourth formant (F24).

After all, the intoxication judgments turned out possible by using the calculated value of F14 and F24. However, the second and third formants are fluctuating depending on the articulatory organs; the 4th formant is not sensitive to them. In this study, therefore, it was cognized that only the slope between the 1st and 4th formants is meaningful.

The process was also required to obtain the formant frequency of the voice signals by using the linear prediction (LPC) algorithm in order to use the formant parameters. Figure 3 shows the significance of the slope of the spectrum, the algorithm proposed in this study, and the formant curve. No ① is the slope of the 1st and 4th formants; ② 1st and 2nd; finally ③ 1st and 2nd.

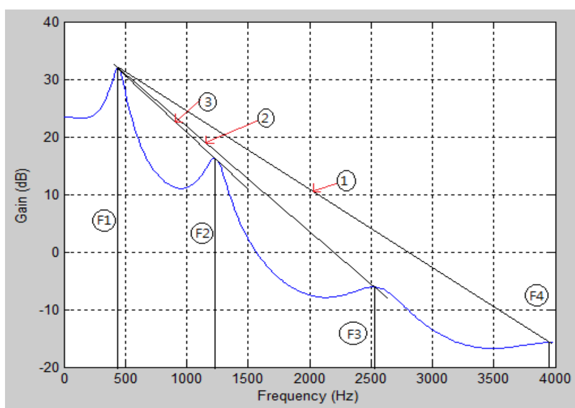

Figure 3. Spectrum slopes

As briefly mentioned earlier, one of the salient features of the voice signal after drinking is that the first pronunciation happens late, and the pronunciation gets less accurate due to the difficulty of the fast pronunciation of the drinker. These phenomena bring the characteristic- the flattening of the spectrum in the frequency domain, which has the same meaning as the fact that the formant curve with a gentle slope in the frequency domain. In other words, drinking makes the slope of the 1st and 4th formant of the frequency characteristics of the speech signal flatter. Therefore, the intoxicated voice can be determined by the extent of the slope of the 1st and 4th formant.

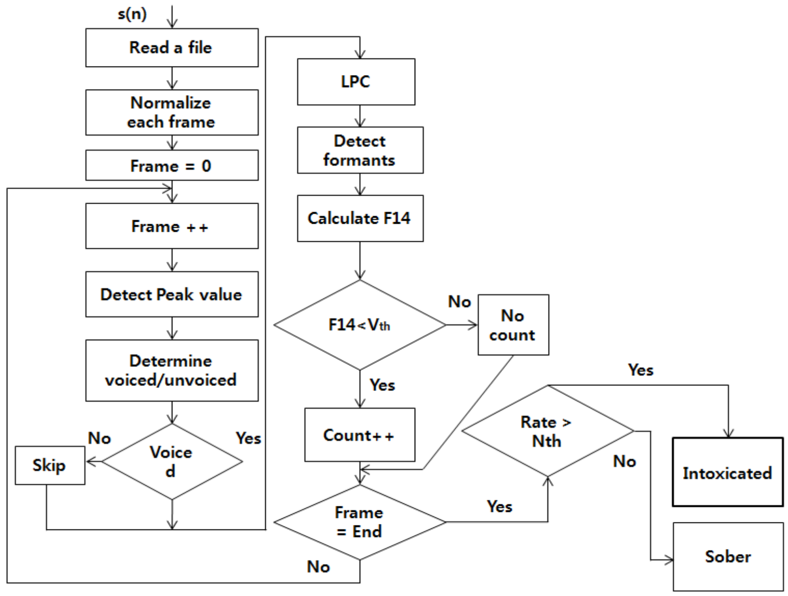

Figure 4. Flowchart of frequency-domain prediction process for intoxication

Figure 4 shows the flowchart for frequency-domain process to predict the extent of drinking. First of all, the process was also required to obtain the formant frequency of the voice signals by using the linear prediction (LPC) algorithm in order to use the formant parameters. F14 was obtained for each frame. If F14 is greater than the threshold, count value is increased. Finally, the Rate is obtained. The Rate means counts divided by the total number of speech frames.

4. Experiments and Results

Samples were taken evenly from each age group between 20's and 60's. 22 standard samples were used when developing the algorithm; same sentence was phonated and a male voice sample without a dialect was selected. The words used for voice sample manuscript were selected from the words difficult to pronounce.

Voice samples were prepared by recording the voice of each person before and after drinking, using mp3. Sound level meter was used when recording in a noisy place. In addition, the degree of alcohol was measured using a breathalyzer before recording; the same person was recorded multiple times to improve the accuracy. The timing of the recording after drinking was determined by subjective feelings and general statistical data, which is generally 1-2 hours after drinking 120% of one's drinking capacity.

The voice files were edited using Adobe Audition program; several sentences pronounced by the same person were edited in constant time intervals and rules. Mat-lab and C programs were used as the simulation program for the development of an algorithm to extract the parameters that determine the intoxication.

As the results of the application of algorithm_1 and algorithm_2 to 22 standard language speaking males with 5% of alcohol concentration, the success rate of intoxication judgment turned out to be 70% in algorithm_1 and to be 71% in algorithm_2.

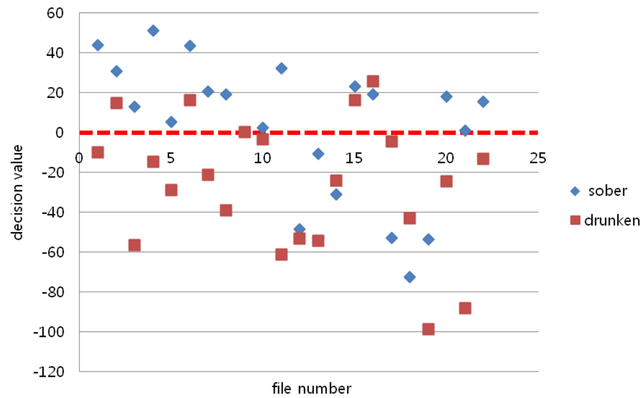

Figure 5. Intoxication prediction results of the voice files

An algorithm was developed in this paper using drunken voice characteristic parameters extracted from time-domain and frequency-domain; each algorithm was weighted depending on the contribution of the accuracy of judgment on drinking. The success rate of intoxication judgment turned out to be 77%.

Figure 5 shows the alcohol concentration and the predicted drinking result for standard sample files. Files from No. 1-22 are voices before drinking; 23-44, after drinking. As shown in the Figure, the voices after drinking were mostly louder than before drinking.

5. Conclusion

In the case of ships, it is difficult to measure blood alcohol level during the sailing. The simplest way to measure the remote ship operator for intoxication is to predict whether the navigator is drunk through communication.

The purpose of the study was on the parameter extraction judging the intoxication by using speech analysis to predict the extent of alcohol drinking through the analysis of the navigator's speech sampled during the communication.

Several important salient features can be found from the two voice signals when compared before and after drinking. In other words, the change in voice energy, change of voice pitch, degraded accuracy of pronunciation, change in speed and duration of pronunciation, early closure of the end of the pronounced sentence, and the change in the pitch and formant are some of the characteristics.

The time-domain parameters used the characteristic that the energy change rate gets smaller because the drinker speaks louder due to the inability of voice control during the phonation from degraded hearing as well as the change in the pronunciation speed and duration after drinking. Secondly, the characteristics of the formant change were used in the frequency-domain, due to the nasalization, dull tongue, and deteriorated hearing ability.

In this study, these characteristics were used to judge the intoxication in the time domain and frequency domain; parameters were extracted to determine the status of drinking. As the result of the application of the algorithm developed using thus extracted weighted parameters, the success rate of intoxication judgment turned out to be 77%.

In future, it is necessary to conduct the research in order to be able to identify the drinking in the continuous signal without having database, that is, speaker's drinking independently from database using these data. In addition this can become the standard to detect the important changes in the speaker's information, and can be expanded as much important parameter by examining the variation in recognizing the speaker.

References

[1] N. -h. Kim, “A Study on intraoral residual alcohol problems and countermeasures in the Drinking measurement process”, Police Studies, no 2, **(2002)**, pp. 5-34.
[2] D. -H. Kyon and M. -J. Bae, “Analysis of Voice Energy after Drinking”, Acoustical Society of Korea and signal processing conference, vol. 29, no. 1, **(2012)**, pp. 105-106.
[3] B.-m. Whang *et al.*, “Analyze the acoustic characteristics of the voice change after drinking and singing”, Speech Sciences, vol. 8, no. 2, **(2001)**, pp. 2-4.
[4] J. -S. Han, “speech signal processing”, Osungmedia, **(2003)**, pp. 12, 60-61.
[5] L. R. Rabiner and R. W. Schafer, “Digital processing of Speech Signal”, **(1978)**, pp. 38-115.
[6] S. -w. Lee, “A Study formant changes in drinking and non-drinking status”, Journal of the Korean Society for Academic Technology, vol. 10, no. 4, **(2009)**, pp. 805-810.
[7] S. -g. Bae, H. -w. Park, G. -r. baek and M. -J. Bae, “A Study on a Drinking Judgement Method of Speech Signal using the Ratio of Harmonic Deviation”, Acoustical Society of Korea and signal processing conference, vol. 29, no. 1, **(2012)**.
[8] G. Baek and M. -J. Bae, “A Study on Judgment of Intoxication State Using Speech”, Communications in Computer and Information Science, vol. 351, Springer, **(2012)**, pp. 277-282.

Authors

Guem-Ran Baek

Address: Soongsil University, 1-1 Sangdo 5 dong DongJak-Ku, Seoul, 156-743, Republic of Korea

Education & Work experience: She received the M.S degree in Electrical Engineering & Electronics from KAIST in 1999. She is currently the Ph.D. degree at Soongsil University and she is the Professor of the Dept. of Computer Science at Soongsil Computer Institute.

Tel: +82-02-828-7337

E-mail: judybaek@daum.net

Myung-Jin Bae

Address: Soongsil University, 1-1 Sangdo 5 dong DongJak-Ku, Seoul, 156-743, Republic of Korea

Education & Work experience: 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 & Tele-communication at Soongsil University.

Tel: +82-02-824-0902

E-mail: mjbae@ssu.ac.kr