International Journal of Multimedia and Ubiquitous Engineering
Vol.8, No.6, (2013), pp.209-216
http:// dx.doi.org/10.14257/ijmue.2013.8.6.21

Speech Sobriety Test Based on Formant Energy Distribution

Chan-Joong Jung[3], Seong-Geon Bae[3], Myung-Sook Kim[2] and Myung-Jin Bae[1]

[2]*Department of English Language and Literature*
[3]*Information and Telecommunication Department, Soongsil University*
Sangdo-Ro 369, Dongjak-Ku, Seoul, 156-743, Korea

cjjung99@hanmail.net, sgbae123@empas.com, kimm@ssu.ac.kr, mjbae@ssu.ac.kr

Abstract

When a person is drunk, lung capacity, lung movement velocity and air pulse change. Drunken people show unique characteristics especially in their speeches with respect to the spectrum area because they tend to speak more forcefully. In other words, the spectrum in the low frequency band area and high frequency band area become relatively flat. This paper suggests that it is possible to develop a sobriety test using speech analysis based on formant energy distribution. The test utilizes the errors in forming the formant envelope between high order linear prediction coding and low order linear prediction coding. About 65% accuracy was obtained in determining whether or not a person is sober by this method.

Keywords: *Sobriety test, linear prediction coding, formant energy distribution, formant envelope*

1. Introduction

Drinking alcoholic beverages may influence someone's behaviors by encouraging them to express more feelings. It makes a person feel happy, sad, excited, and relaxed. However, driving or sailing after drinking alcohol causes many problems. Therefore, one should not drink before driving or sailing and it has been regulated by laws in most countries. Still, the accident rate is very high and the statistic data prove that many of them were caused when sailors/mates were alcoholically intoxicated. Table 1 is the annual report for drunk sailing cases published by the Road Traffic Authority of Korea. As shown in the table, the number of cases is somewhat consistent. The age range of drunken sailors is between 20s and 40s [1].

Table 1. Numbers of drunk sailing cases from 2007 to 2011

Year	Number of driving-under-influence cases
2007	28,416
2008	26,873
2009	28,207
2010	28,641
2011	28,461

[1] Corresponding author : Myung-Jin Bae (*mjbae@ssu.ac.kr*) Soongsil University, Sangdo-Ro 369, DongJak-Ku, Seoul, 156-743, Republic of Korea.

ISSN: 1975-0080 IJMUE
Copyright ⓒ 2013 SERSC

Although education and publicizing on safety driving have been continuously conducted in a variety of fields, the awareness on safety driving has not reached to an adequate level yet. The recent trend is reinforcing punishment by revising related laws. Like safe driving on land, safe sailing is essential on the sea. Safety sailing means safely operating all types of boats (including floating aircraft and wing-in-ground effect ship) which are used for moving on the water [2].

As far as accidents related to ships are concerned, there are two representative characteristics on the Korean sea in particular. Firstly, once the accident occurs, large damage is likely suffered although the accident could have been prevented in advance. In other words, most of the accidents were caused by human errors. Secondly, there were many cases the initial accident caused the secondary damages including contaminating the water, which could also have been prevented at the beginning stage of the accident. They were also caused due to human errors [3]. Therefore, we could conclude that the accidents on the sea should be prevented from the very beginning.

Accidents on the sea can be classified into several types such as collision and sinking, but the major cause of them is mostly safety negligence. To prevent safety negligence, drunk sailing in particular, the police implemented a policy for measuring blood alcohol level before sailing. According to the 2011 report by Korea Coast Guard as shown in Table 2, the number of cases for measuring blood alcohol level on the sea is decreasing while number of crackdown cases for drunken sailing is not, keeping its consistent tendency.

Table 2. Yearly report for measuring blood alcohol level and drunken sailing [8]

Year	Number of cases for measuring blood alcohol	Number of crackdown cases for drunken sailing
2010	16,357	122
2009	20,965	148
2008	56,420	142
2007	56,420	184
2006	64,921	127

If measuring blood alcohol level is difficult to implement on the sea, we should come up with the alternative. Since it is usual to communicate frequently between the ship and the land using voice, we suggest that a sobriety test should be developed using speech analysis provided that people show unique characteristics in their speech when they are alcoholically intoxicated.

Humans make sound by vibrating vocal cords at a constant speed and vocalize words by changing the shape of vocal tract, changing the place of contact points, and changing the manner of producing sounds, with respect to pharyngeal cavity, oral cavity and nasal cavity, using active articulators including tongue, jaw and lips. With the help of all those articulators, humans are able to produce various consonants and vowels [4].

When a person drinks alcoholic beverages, physical changes occur. Most of all, his/her voice quality becomes different. First, when a person drinks, alcohol gets into blood and makes blood circulation smooth. Since nose is made of mucous membrane, the membrane is expanded as blood supply increases. Then the nose is blocked and nasal sound appears. A characteristic feature of nasal sound is that the formant bandwidth increases while the sound source made by vocal cord vibration becomes weaker as it passes through the vocal tract [4].

 Copyright ⓒ 2013 SERSC

Second, the acoustic features of nasal sound can be characterized as follows. Since nostril becomes narrower, nasal sound is formed at the frequency band with low energy such as the 1st formant (or F1). On the spectrogram, strong energy appears at 300Hz band. It also affects the vowels before and after the nasal sound. It decreases the energy of the 1st formant. The interval of formant below 800Hz is short while the formant bandwidth is widen [5].

Third, when a person drinks alcohol, the person talks in a loud voice. The reason is that alcohol (ethanol) works on the central nerve system of human body and restricts the nerve activity of human brain. In other words, alcoholic drink decreases performance readiness in physical movement, rational judgment, self-control, concentration, cognitive ability and situation judgment capability.

2. Speech Sobriety Test based on Formant Energy Distribution

It is difficult to directly measure alcohol level in the blood vessels of sailors while a ship is sailing. Therefore, this paper suggests a non-contact sobriety test by analyzing voice sample. First, normalization will be done by frame and each frame is reduced to 70% level. Since this study uses the slope and angle between formants, we will pay attention only on voiced sound section.

This study also suggests a new method for differentiating voiced sound and voiceless sound coming out of voice signals. The number of peak values in voice signal wave patterns will be a crucial factor in determining whether or not a certain sound is voice or voiceless. If the number of peak value is too small or too big, it is considered either a silent sound section or voiceless sound section; therefore, it will be excluded.

First, the formant frequency of voice signals will be obtained by linear prediction coding (LPC). There are two ways in getting formant frequency. One is the root solving method, which calculates the root of inverse filter. The other is the peak picking method, which calculates interpolated spectrum of inverse filter by FFT and finds the peak value of a spectrum.[6]

It is possible to get all formant values by using the root solving method. In the root solving method, formant bandwidth on complex root 'z' (B) and formant frequency (F) are obtained by converting plane 's' to plane 'z'.[7]

$$z = e^{st} \tag{1}$$

Here, $s = -\pi B \pm j2\pi F$; and,

$z = R_e(z) \pm jI_m(z)$ is defined as both a real number in part and an imaginary number in part of complex root. Formula (2) is formant frequency while formula (3) is formant bandwidth.

$$F = (f_s / 2\pi)\tan^{-1}[\mathrm{I}_m(z)/R_e(z)] \qquad (Hz) \tag{2}$$

$$B = -(f_s / \pi) In|z| \qquad (Hz) \tag{3}$$

The acoustic features of the 1st, 2nd and 3rd formants can be characterized as follows:

The 1st formant is related to the opening extent of mouth. When mouth is more open, the 1st formant is higher. Accordingly, the 1st formant is lower when mouth is less open. The 1st formant is low at high vowels and high at low vowels. The 1st formant becomes low when the narrowing point appears at the front part of vocal tract, while it becomes high when the narrowing point appears at the back part of vocal tract.

Copyright ⓒ 2013 SERSC

The 2nd formant is high at front vowels and low at back vowels. As the narrowing point occurs at the more front part of glottis and palate, the 2nd format gets higher. As the narrowing occurs nearer to lip or soft palate, the 2nd format gets lower.

The 3rd formant gets higher when the narrowing point occurs nearer to glottis, soft palate and teeth ridge. It gets low when the narrowing point occurs at the upper side of lip, soft palate and pharynx.

Figure 1 is the block diagram of the speech sobriety test suggested in this paper. In the first, voiced/unvoiced detection. If it is a voiced sound, it passes through low order (LPC Order: 4) and high order (LPC order: 12).

Using formula 4, 1st formant and 2nd formant of the slope obtained. Prediction error (PE) is calculated using formula 5. The slope of the formant is less than the threshold value and Prediction error (PE) is less than the threshold value, it will determine that a person is drunk.

$$Factor = \frac{Gain_2 - Gain_1}{F_2 \max - F_1 \max} \tag{4}$$

Here,

$F_2 \max$: Peak value of the 2nd formant

$F_1 \max$: Peak value of the 1st formant

$Gain_2$: The 2nd formant gain

$Gain_1$: The 1st formant gain

Prediction Error (PE) = | LPC Filter (Order :12) - LPC Filter (Order :4) |^2 *100 (5)

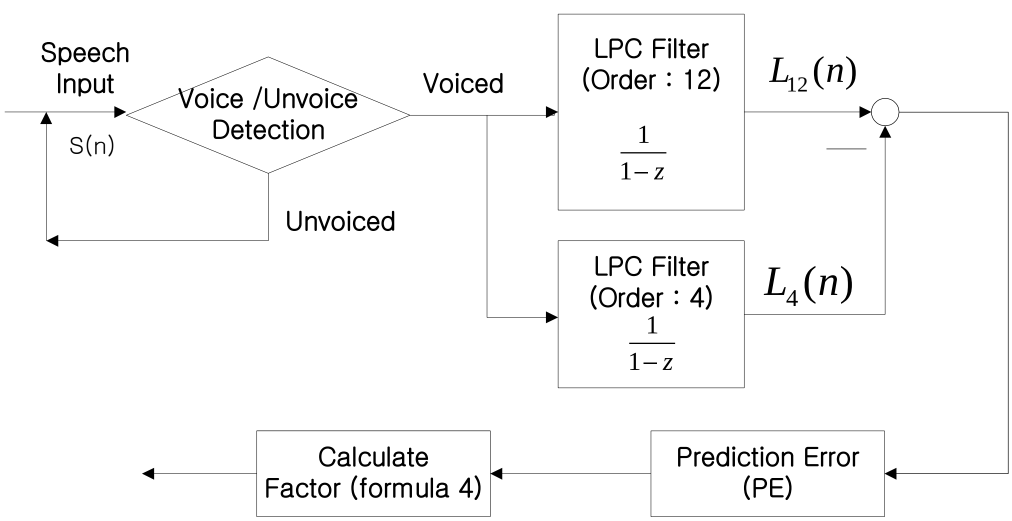

Figure 1. The block diagram for speech sobriety test

 Copyright ⓒ 2013 SERSC

Figure 2 is the envelope obtained by high order LPC and the envelope obtained by low order LPC. If a person is drunk, conspicuous peaks appear at formant locations and the slopes from the 1st formant up to the 4th formant are relatively flat.

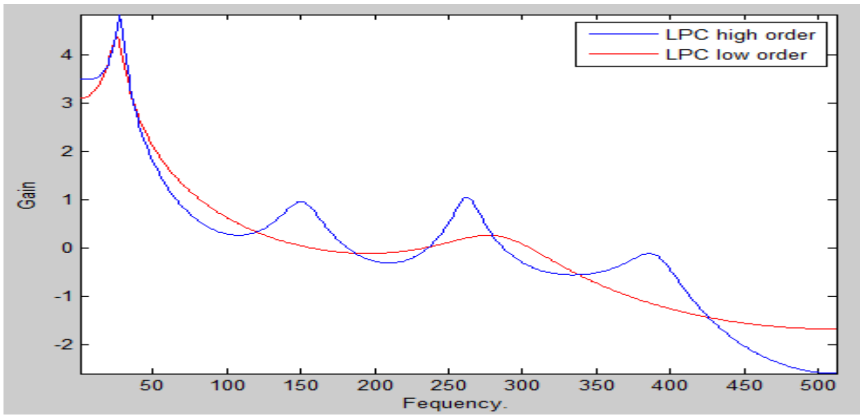

Figure 2. High Order LPC / Low Order LPC

When the spectrum shows relatively flat slopes between the 1st formant and the 4th formant, there will be less error in high order LPC, which follows linear predictive relatively well. However, forecast error becomes bigger in low order LPC. The predictive error is obtained by subtracting low order LPC from high order LPC and it is used for judging whether or not a person is drunk.

3. Test and Result

Since communication with ship is usually conducted through radio with various noises from the environment, mp3 player is used in the experiment for this study because it has a bit poorer sound quality than a high performance audio recorder. The voice samples were taken from a wide range of sailors from 20s to 60s. Sampling rate is 8kHz.

Voice recording was done when two hours passed from the time of drinking alcohol. The sailors read a paragraph which took about 5 minutes to read and their readings were recorded.

Table 3 is the result of the experiment using the method for speech sobriety test suggested in this paper. When a person is drunk, he/she shows different spectrum in their speech. The slopes from the 1st formant to the 4th formant become relatively flat.

High order linear filter follows the forecast relatively well; however, the forecast error of low order linear filter becomes bigger. Utilizing this characteristic feature of speech, formula (4) is used for speech sobriety test.

Figure 3 is the flow chart of algorithm suggested in this paper. When a voice signal comes in, it passes through low order (LPC Order: 4) and high order (LPC order: 12). If the value obtained by formula (4) is smaller than the threshold value, it is judged that the person is drunk. If it is larger than the threshold value, it is judge that the person is sober.

Copyright ⓒ 2013 SERSC

Table 3. Decision of sobriety

Name of file	Factor (Equation 4) (Average)	PE	Decision
Sm201	0.3312	8	Drunk
Sm202	0.3301	7	Drunk
Sm203	0.2910	3	Drunk
Sm204	0.2934	5	Drunk
Sm205	0.2656	4	Drunk
Sm206	0.2554	1	Drunk
Sm207	0.2588	7	Drunk
Sm208	0.3509	6	Drunk
Sm209	0.2680	5	Drunk
Sm210	0.3258	7	Drunk
Sm211	0.3521	8	Drunk
Sm212	0.3832	21	Sober
Sm213	0.3365	5	Drunk
Sm214	0.3227	4	Drunk
Sm215	0.3195	2	Drunk
Sm216	0.3633	15	Sober
Sm217	0.3615	13	Sober
Sm218	0.3618	16	Sober
Sm219	0.3728	19	Sober
Sm220	0.3872	21	Sober
Sm221	0.3648	17	Sober
Sm222	0.3662	16	Sober
Sm223	0.3456	6	Drunk
Sm224	0.3556	7	Drunk
Sm225	0.3558	7	Drunk
Sm226	0.3797	22	Sober
Sm227	0.3936	25	Sober

Copyright ⓒ 2013 SERSC

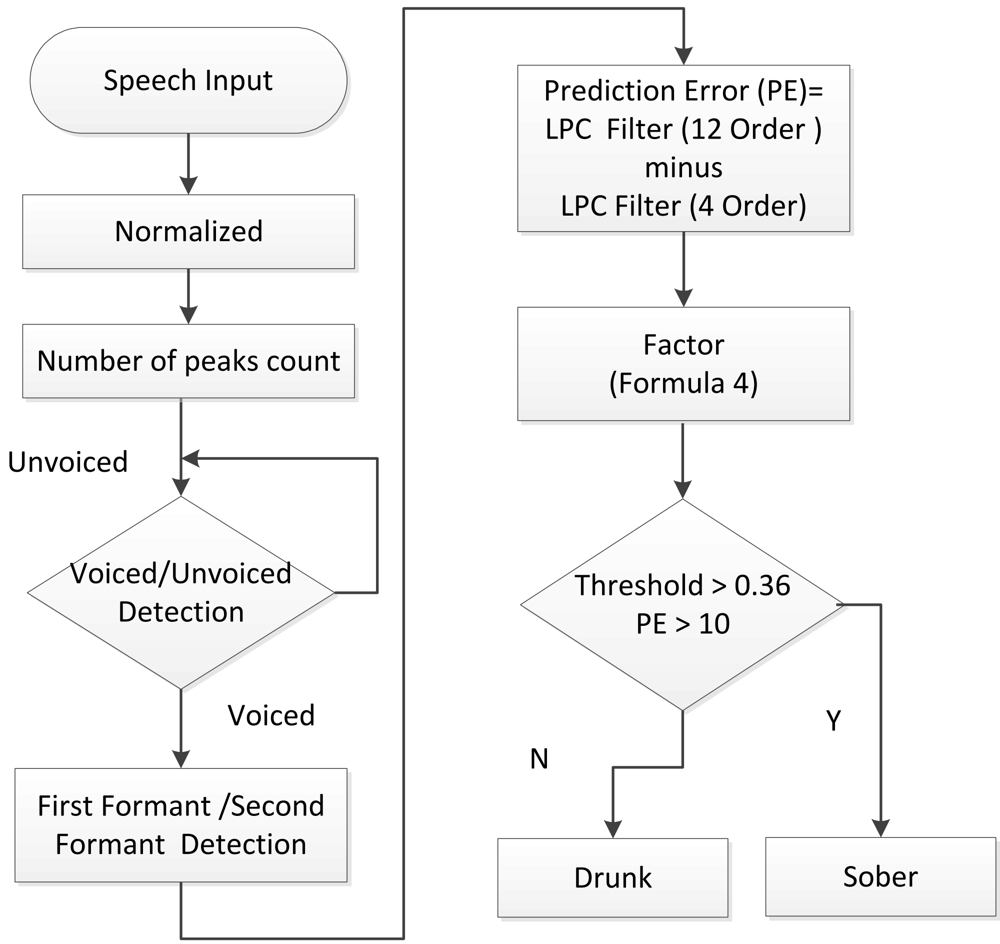

Figure 3. Flow chart of speech sobriety test

4. Conclusion

Safe sailing on sea requires high level of concentration; however, accidents happen because sailors cannot keep concentrated due to various reasons. Drunk sailing can be regarded as one of the most popular causes for accidents on the sea, because alcoholic drinks decrease performance readiness in physical movement, rational judgment, concentration and cognitive ability.

Accordingly, this study suggested a sobriety test method which utilizes human speech voices. The purpose of the study was on developing a parameter to be used for judging whether or not a person is alcoholically intoxicated using speech analysis based on formant energy distribution.

When a person is drunk, the spectrum shows relatively flat slopes; therefore, there will be less error in high order LPC, which follows linear prediction relatively well. However, prediction error becomes bigger in low order LPC. The prediction error is obtained by subtracting low order LPC from high order LPC. This prediction error is used for sobriety test. The speech sobriety test suggested in this paper yielded 65% accuracy.

References

[1] National Statistical Office, http://www.index.go.kr/.
[2] I. -H. Kim, "Korean Marine Traffic Law", (3rd Edition), (**2011)**.
[3] D. J. Kim and S. Y. Kwak, "Evaluation of Human Factors in Ship Accidents in the Domestic Sea", Journal of the Ergonomics Society of Korea, vol. 30, no. 1, (**2011)** February, pp. 87-98.
[4] B. Yang, "Measurement of the vocal tract area of vowels By MRI and their synthesis by area variation", Korean Journal of Speech Sciences, vol. 4, no. 1, (**1998)**.
[5] S. Choi and C. -J. Seong, "Long Term Average Spectral Analysis for Acoustical Discrimination of Korean Nasal Consonants", Journal of the Malsori and Speech Science, (**2006**), pp. 67-84.
[6] R. C. Snell and F. Miliazzo, "Formant Location From LPC Analysis Data", IEEE Transactions on Speech and Audio Processing, vol. 1, no. 2, (**1993)** April.
[7] T. -W. Kim, "Korean Speech Recognition Preprocessing System Using Formant Frequency", InHa University Master Thesis, (**2000**).
[8] Korea Coast Guard 2011 White Paper.

Authors

Chan-Joong JUNG

M.S. in Electronic Engineering, University of Incheon. 2000.
ABD in Electronic Engineering, Soongsil University.
Areas of Specialization: Speech Signal Processing, Noise Reduction.

Seong-geun BAE

M.S. in Electronic Engineering, Konkuk University. 1995
ABD in Electronic Engineering, Soongsil University
Professor of Electronics and Telecommunications, Daelim University
Areas of Specialization: Speech Analysis, Speech Synthesis, Speech Coding, Acoustic Analysis.

Myung-Sook Kim

Ph. D. in English
Associate Professor of English, Soongsil University
Areas of Specialization: English Phonetics.

Myung-Jin BAE

Ph.D. in Electronic Engineering
Professor of Information and Communication and Electronic Engineering, Soongsil University

 Copyright ⓒ 2013 SERSC