Journal of Engineering and Applied Sciences 13 (5): 1093-1097, 2018
ISSN: 1816-949X

On a Study on Drinking Judgement Using Phonation Characteristics in Speech

Won-Hee Lee, Kwang-Bock You and Myung-Jin Bae
Department of Information and Telecommunication, Soongsil University, Seoul, Korea

Abstract: In this study, speech characteristics from before and after alcohol intoxication have been comparatively analyzed through speech analysis to obtain the degree of intoxication along with its parameters. The speech characteristics of an alcohol intoxicated person such as pitch and formant and changing of sound levels have been studied but the speech characteristics before and after alcohol intoxication cannot be analyzed because of the sensitiveness of environmental changes. Therefore, we need to find more precise parameters, especially, phonation threshold pressure, the high levels of which result in dehydration of the laryngeal mucous membrane after alcohol intoxication. Speech after alcohol intoxication is with low lung pressure and decreasing lung capacity. Thus, we study the speech characteristics from before and after alcohol intoxication using speech-rate in speech signal and lung capacity.

Key words: Multiple sequence alignment, pairwise sequence identity, sequence selection treatment, transmembrane protein, transmembrane segment, alcohol intoxication

INTRODUCTION

Sequence the voice sobriety test has many advantages compared to traditional sobriety tests such as field sobriety test (testing whether or not a person can walk straight on the line), blood test (taking blood sample to measure alcohol level in blood) or breathalyzer test (breathing into the breathalyzer to measure alcohol level in breaths). Firstly, it needs no physical contact between the tester and testee (one of the compulsory requirements for most traditional tests), since, the voice can be wired through communication devices if necessary. Secondly, the test can be done without arousing the testee's awareness. Just ordinary talking to any type of microphone (or voice processing device) makes it possible to administer the test. Lastly, the test can be efficiently and promptly implemented anytime and anywhere by anyone, since, the voice processing program automatically determines the degree of intoxication and delivers the results on the spot to relevant parties in question.

Especially in case of sailing under the influence of alcohol which is very common among fishermen especially in the Korean sea, implementation of the sobriety test based on speech analysis through wireless communication channels can be the most feasible and the most effective way to prevent accidents without needing physical contact.

MATERIALS AND METHODS

Human beings produce (or articulate) desired speech sounds by utilizing speech organs or the vocal tract (the collective term of the pharyngeal cavity, the oral cavity and the nasal cavity) usually while breathing out air from the lung. We produce various sounds in diverse manners of articulation such as voiced and voiceless sounds, plosive and fricative sounds as well as various sounds from different places of articulation inside of our vocal tract such as labial sound around lips and alveolar sound around alveolar ridge.

The voice production model used in this study is illustrated in Fig. 1. In this modeling, Gaussian noise is considered as one of input source for the processing in addition to fundamental frequency from vocal cords, since, the human voice as well as telecommunication channels and computer networks evidently produce turbulent sound and noise. Detailed explanation for the voice production modeling used in this study can be found by Bae and Bae (2016).

When we consume alcoholic beverages, the alcohol contacts the vocal tract first before getting absorbed into the esophagus and the gastrointestinal tract. Continuous intakes of hard liquor such as whiskey, soju (one of the most popular hard liquors in Korea) and vodka may act as strong stimuli to mucous membrane which surrounds the esophagus or the top layer (epidermis) and

Corresponding Author: Myung-Jin Bae, Department of Information and Telecommunication, Soongsil University, 369 Sangdo-ro Dongjak-gu, 156-743 Seoul, Korea

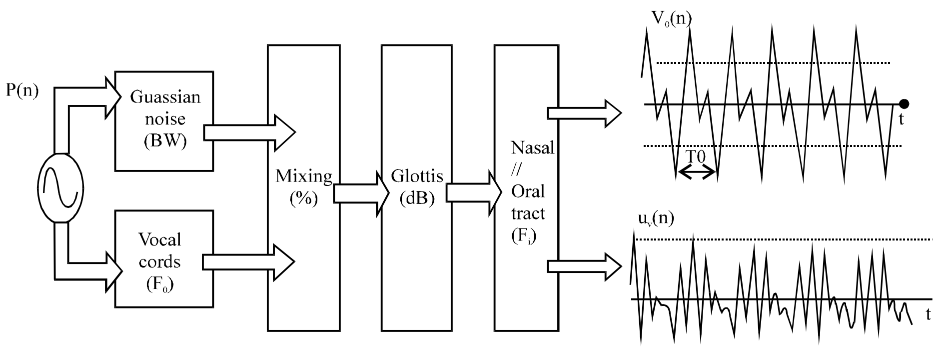

Fig. 1: Voice production modeling

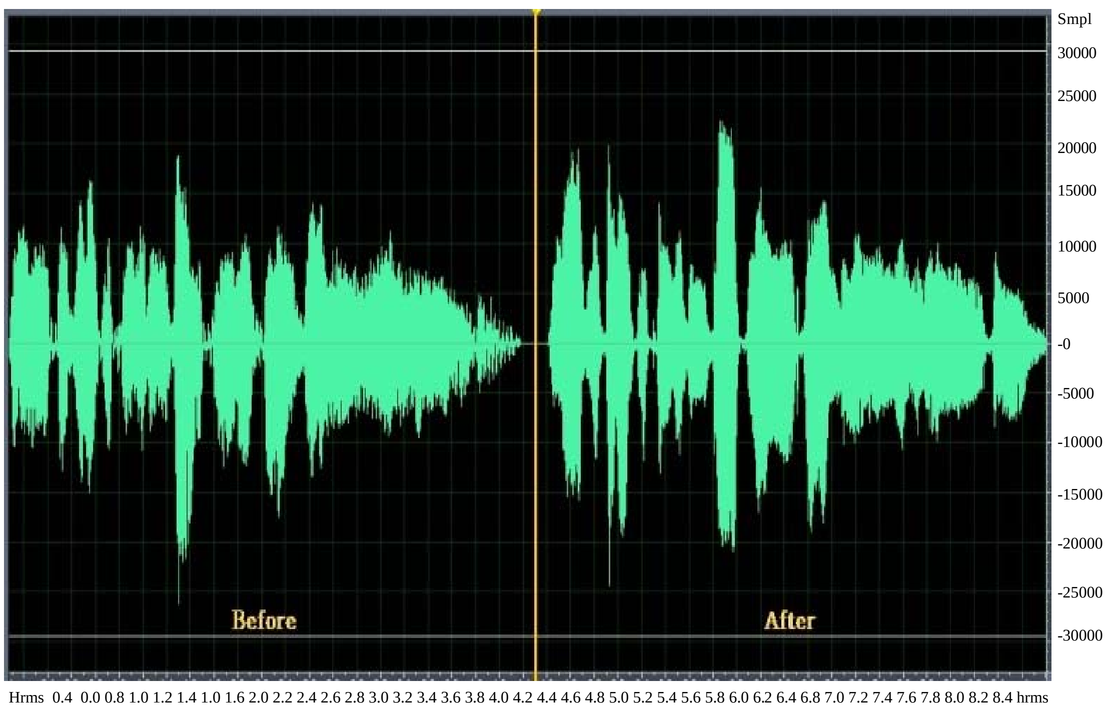

Fig. 2: Waveforms of speech signals before and after drinking

cause dehydration in these or gans including the vocal tract. Needless to say, most alcoholic beverages have diuretic effects aggravating dehydration.

As dehydration occurs in these organs, the flexibility and elasticity of the vocal cords may be reduced and thus we may experience incomplete opening/closing of the vocal cords, especially when sub-glottal pressure is applied. Eventually the clarity of pronunciation becomes degraded (Ajgou *et al*., 2014). The incomplete opening/closing of the glottis boosts the threshold pressure for phonation, so that, the speaker needs more power to initiate and maintain the phonation and therefore the energy (or amplitude) of the voice increases. The pronunciation (or articulation of sounds) starts late and it gets less accurate due to the difficulty of uttering the fast speech. The incomplete opening/closing of the mouth may also cause the decrease in clarity of pronunciation. With decreased clarity of pronunciation and increased amplitude of the voice, the pitch can play an important role as a parameter to detect changes in voice when intoxicated. Nasalization of the voice can also be detected from the voice intoxicated (Bae and Bae, 2016; Kyon and Bae, 2012).

Changes in voice when intoxicated are compared in Fig. 2 showing waveforms and Fig. 3 showing spectrums of voices before and after drinking. In Fig. 2, vocal

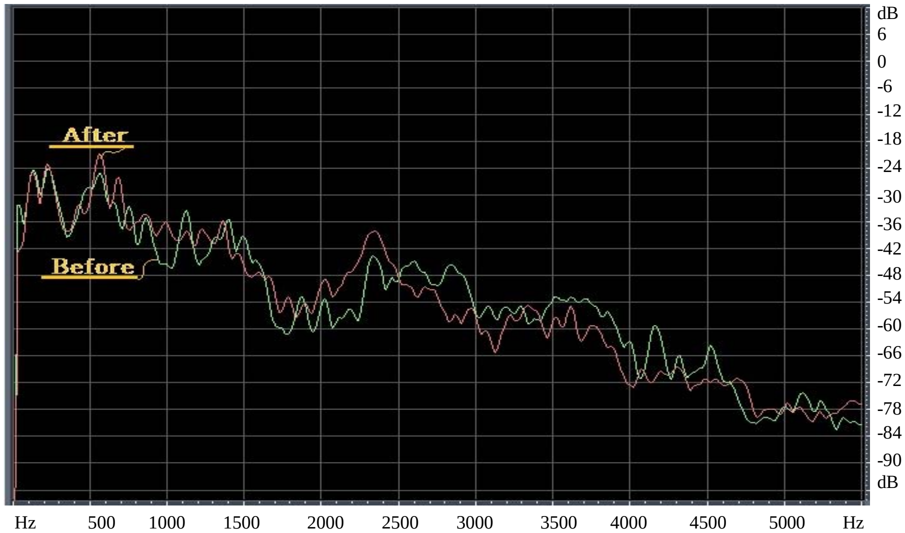

Fig. 3: Spectrums of voices before and after drinking

differences between before and after drinking can be found in irregular shapes of waveforms and in their amplitude. In Fig. 3, the first formants detected in voices before and after drinking appear similar to each other but the second formants look quite different from one another, showing some degree of nasalization of voice. Other differences can also be detected in voices between before and after drinking (Kyon and Bae, 2012). All these differences and others in voice before and after drinking have been carefully and thoroughly considered to develop parameters as well as algorithms for voice processing in the study.

RESULTS AND DISCUSSION

Parameters to be used in voice processing for sobriety test can be observed mainly from two basic domains: the time domain and the frequency domain. There are of course, additional factors to be considered as parameters for the voice sobriety test which will be dealt with in transform domain, a hybrid domain of time and frequency.

Algorithms from time domain include equations to calculate some popular processes known as ZCR (Zero Crossing Rate), Autocorrelation and AMDF (Average Magnitude Different Function). The detailed algorithms can be found by Abdulla (2002) and Bae and Bae (2016).

In the frequency domain, we also get some unique characteristics of the intoxicated voice including nasalization of the voice caused by drinking alcohol (Bae and Bae, 2016; Jung *et al*., 2013). The nasalized voice tends to have wider formant bandwidth. The speaker produces louder voice while opening his/her mouth wide and thus the lowest formant appears differently from that of the sober voice. The discrepancy in the position of the tongue (being positioned further back than usual) also affects the size of the second formant. Therefore, the relationship between the slope of the 1st and 4th formant (F1-4) and the slope of the 2nd and 4th formant (F2-4) can be captured in the expression of ‘F1-4<F2-4’ to characterize intoxicated voice. The slopes of these two cases are illustrated in Fig. 4.

The algorithms from frequency domain will cover processing like FFT (Fast Fourier Transform) and Cepstral analysis. These algorithms developed for utilizing time domain as well as frequency domain are weighted based on the accuracy of judgment for the intoxicated voice. The specific algorithms are explained by Baek and Bae (2013).

There is something else, we should consider in order to substantiate a more accurate way to process the voice. We may call it transform domain, a hybrid domain of time and frequency domain. There are 2 parameters accommodated into the voice sobriety test: the LPC (Linear Predictive Coefficient) with error prediction and formant energy distribution (Jung *et al*., 2013).

An experiment was conducted to justify the assumption that human voice may be processed to judge whether or not a speaker is intoxicated based on voice qualities or more specifically changes in voice qualities. Voice samples were taken from ten people in each age group between 20 and 60’ sec. Out of 50 voice samples, 22 standard samples were used as control samples when

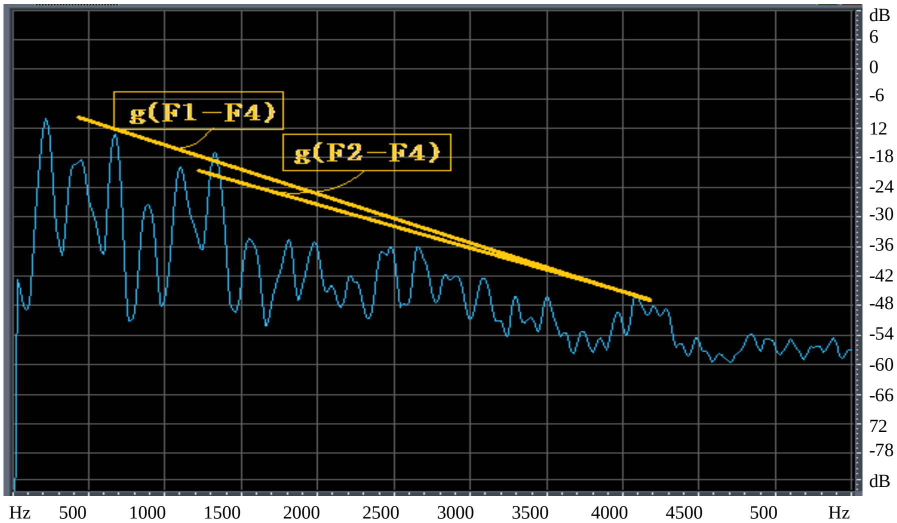

Fig. 4: The formant slopes difference for F1-4 and F2-4

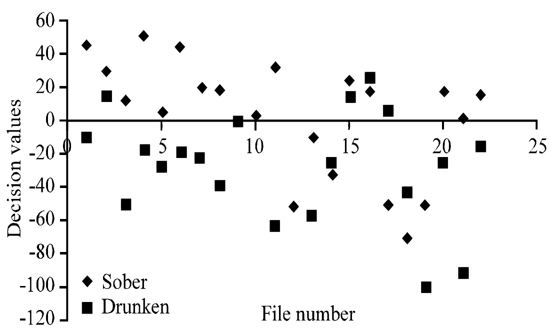

Fig. 5: Judgement result for intoxication

developing the algorithms for standard samples, the same sentence was phonated and a male voice without any dialectal or regional accent was selected as the sample voice. The words and sentences used in voice sample were selected with high considerations for particularly difficult words to pronounce and words with high frequency among fishermen.

The result is illustrated in Fig. 5. Notice that the voices where the speaker had been drinking are generally located under the standard deviation zero line and the voices where the speaker is sober generally stay above the zero line. However as illustrated in Fig. 5, some samples are found against the expectation, producing a wrong judgment.

We should point out that applying the parameters and algorithms to phonation voice processing in this case can be done using 2 methods: speaker-dependent and speaker-independent. The speaker-dependent method compares voices from the same speaker before and after consuming alcohol. On the other hand, the speaker-independent method processes the voice of the speaker without comparing the voices between before and after drinking.

The experiment result shows that, the rate for accurate judgment of intoxication is 82% when using the speaker-dependent method and 72% when using the speaker-independent method, thus, yielding an average of 77% judgment rate. Although, the judgment rate is somewhat acceptable for the speaker-dependent processing method, we will further pursue the speaker-independent processing method for a few reasons. Firstly, the judgment rate is comparatively better than that of speaker-dependent method as mentioned earlier. Secondly, the speaker-independent method is more prospective in application to the real world, since, it will not require before and after sample recordings in order to administer the test fully.

CONCLUSION

This study certainly confirms that the voice sobriety test is an effective and reliable way to be utilized for many practical purposes. Not only the captains at sea but also operators of public transportation such as pilots of commercial airplanes, bus and taxi drivers and train operators can be subjects in the voice sobriety test to ensure the safety and well-being of people. Therefore, we need to find more precise parameters, especially, Phonation threshold pressure, the high levels of which result in dehydration of the laryngeal mucous membrane after alcohol intoxication. Speech after alcohol intoxication is with low lung pressure and decreasing lung capacity. Thus, we study the speech characteristics from before and after alcohol intoxication using speech-rate in speech signal and lung capacity.

REFERENCES

Abdulla, W.H., 2002. Auditory based Feature Vectors for Speech Recognition Systems. In: Advances in Communications and Software Technologies, Mastorakis, N.E. and V.V. Kluev (Eds.). WSEAS Press, Saskatchewan, Canada, pp: 231-236.

Ajgou, R., S. Sbaa, S. Ghendir, A. Chamsa and A. Talebahmed, 2014. Novel detection algorithm of speech activity and the impact of speech codecs on remote speaker recognition system. WSEAS. Trans. Signal Process., 10: 309-319.

Bae, S.G. and M.J. Bae, 2016. Reducing errors of judgement of intoxication in overloaded speech signal. Intl. J. Eng. Technol., 8: 219-224.

Baek, G. and M. Bae, 2013. A study on voice sobriety test algorithm in a time-frequency domain. Intl. J. Multimedia Ubiquitous Eng., 8: 395-402.

Jung, C.J., S.G. Bae, M.S. Kim and M.J. Bae, 2013. Speech sobriety test based on formant energy distribution. Intl. J. Multimedia Ubiquitous Eng., 8: 209-216.

Kyon, D.H. and M.J. Bae, 2012. Analysis of voice energy after drinking. Acoust. Soc. Korea Signal Process. Conf., 29: 105-106.