International Journal of Engineering Research and Technology.
ISSN 0974-3154 Volume 11, Number 9 (2018), pp. 1463-1470

http://www.irphouse.com

A Study on Identification Rate Difference of Sound Color Marker According to Bandwidth-limited of Voice Signal

Bong-Young Kim
[1]*Department of Information and telecommunication Engineering*
Soong-sil University, Seoul, Korea

Myung-Jin Bae*
[2]*Department of Information and telecommunication Engineering*
Soong-sil University, Seoul, Korea

ABSTRACT

There are many ways in which a person can identify a person. Biometric authentication uses a wide variety of methods, including fingerprint, voice, face, iris, the lines of the palm, and vein distribution. Of the various authentication methods, the voice authentication method is a very effective authentication method in terms of security, convenience, and economy. Because voice signals contain a lot of information, the results of differentiating personality will vary depending on how bandwidth-limited is done. In this study, the identification rate difference according to the bandwidth-limited of the voice signal was confirmed by using a sound color marker that expresses the voice signal into seven colors. As a result of the experiment, it was possible to identify the same person or others in various cases of bandwidth-limited voice signal only by comparison of sound color marker. Also, it can be seen that, even if the bandwidth is limited to various forms, all of them have individual characteristics. As a result, the sound color marker reflects the bandwidth-limited individual characteristics, so that even if bandwidth-limited, identification of the same person or other person is possible.

Keywords: Biometric Authentication, Voice Signal, Bandwidth-limited, Sound Color Marker, Identification Rate

* *Corresponding author: Myung-Jin Bae, Professor.

I. INTRODUCTION

There are many ways in which a person can identify a person. Usually, when we meet the other person, we can check who is the other person through face shape, height, body shape, etc., and even if we do not look with the naked eye, we can distinguish who is the other person by using voice, footsteps and body odor. Usually, when we meet the other person, we can check who is the other person through face shape, height, body shape, etc., and even if we do not look with the naked eye, we can distinguish who is the other person by using voice, footsteps and body odor. Identify the other person using the physical and habitual characteristics of each person. Biometric Authentication can be used to identify who is Biometrics. Biometric authentication uses a wide variety of methods, including fingerprint, voice, face, iris, the lines of the palm, and vein distribution. Multiple Biometric Authentication, which improves accuracy and security by using several methods together, is also being developed. Of the various authentication methods, the voice authentication method is a very effective authentication method in terms of security, convenience, and economy. In the case of face, the lines of the palm, and iris, there is a fear of hacking and duplication of Biometrics using video signal processing technology and 3D printing technology. However, voice authentication is very safe for lost, stolen, duplicated, etc., and it is possible to maximize security because it can present sentences to be authenticated. In the case of various biometric authentication methods, the device for collecting Biometrics is very expensive. In case of voice authentication, however, there is already a microphone or voice processing device installed in every smart device or computer to process the voice signal. Therefore, voice authentication is a biometric authentication method that can be introduced immediately at low cost [1,2].

The range of frequencies that a person can perceive as an ear is known as 20 to 20,000Hz. When transmitting conversation contents through electronic means such as telephone, voice is transmitted at 4000 Hz or less. However, when the voice is transmitted at 4,000Hz or less, it sounds very different from the actual voice, and sometimes it is difficult to clearly distinguish the voice of the other party. This means that even in a band of 4,000Hz or more of voice, information indicating the characteristics of speech and voice exists. When the band is limited to 4000Hz or less, the characteristics are lost and the speech power of speech is lower than the actual voice. Therefore, depending on how you limit the bandwidth of the voice signal, the result of identifying each individual will vary [3,4].

In this study, the identification rate difference according to the bandwidth-limited of the voice signal was confirmed by using a sound color marker that expresses the voice signal into seven colors. Chapter 2 describes the characteristics of the voices and the Sound Color Marker. Chapter 3 describes experiments and results, and Chapter 4 conclude.

II. CHARACTERISTICS OF VOICE AND SOUND COLOR MARKER

Characteristics of Voice : For human voices, phonological characteristics are usually expressed using fundamental frequency(F0) and formant frequency(F1 ~ F5). The fundamental frequency is the period in which the vocal cords are shaken by the air boosted from the lung, and the range of the fundamental frequency that is generated by the human larynx structure is determined. The fundamental frequency goes up or down depending on factors such as the speaker's strength, intonation, and emotion. The formant frequency is determined by the fundamental frequency of the speaker and the structure of the vocal tract. Depending on the fundamental frequency and the length of the vocal tract, a resonant frequency is generated, which is classified as 1'st formant(F1) to 5'th formant(F5) according to the order of the resonant frequencies. When a person speaks, the structure of a vocal organ changes, and the distribution of each formant frequency changes. Such a change determines the phonological properties of the phoneme. In general for voyelles, the phonological properties of phonemes are determined mainly by F1, F2. And F3, F4, F5 mainly contain human characteristics. However, in the case of fricatives, the formant is very complicated and contains many phonological information in F3, F4 and F5. Therefore, the frequency range of the high region also has phonological characteristics such as fricatives as well as personality [5,6,7].

Sound Color Marker : The auditory organ that perceives sound is recognized as a log scale with respect to the height or size of sound. When the frequency is increased by 10 times, it is recognized as twice the height. When the frequency is increased by 100 times, it is recognized as being three times as high. The various scales developed using the characteristics of this hearing are used only for the specific experts concerned with sound. The Sound Color Marker is a way of expressing the scale of sound to people so that they can understand it intuitively. Human visual characteristics, like auditory characteristics, are perceived as log scale for the color or intensity of light. Using this similarity of visual and auditory characteristics, the Sound Color Marker is a representation of the hearing response in a seven-color ratio similar to the sense of light. Through these visualized expressions, even if there is no knowledge of the hearing, the scale can be perceived intuitively and the complex parameters of the voice signal can be simplified to seven [8,9].

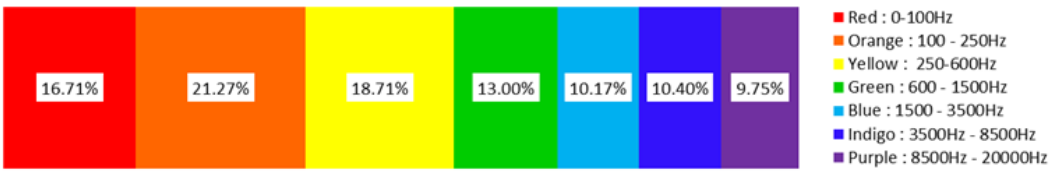

Fig 1. The process of calculating SDNN from heart rate measurement

III. EXPERIMENTS AND RESULTS

In the experiment, we tried to compare the identification rate of each individual by using Sound Color Marker by bandwidth-limited voice data of 4KHz or less, 20KHz or less, 4KHz ~ 20KHz respectively. We collected the sound sources from 10 different men and women by reading "There lived mother pigs and three little pigs together in deep mountains" 3 times each. The collected sound sources were bandwidth-limited in various ways and then converted into Sound Color Markers, respectively. The generated Sound Color Markers were compared with each other to confirm whether the same voice was identified through similarity check. The sound source was collected using Samsung Galaxy Note 4, which is a Samsung smartphone, and sampled at 44100 Hz and 16 bit quantized. The software used was Audition CC and Cool Edit Pro 2.1. The sound sources are called M-1, M-2, W-1, and W-2 according to the gender and order of the person who collected the sound source. We collected the sound sources three times for each 10 persons, 5 for each man and woman. They were divided into A to C according to the order of the sound sources collected from the same person. Each sound source is called M-1-A. For the 30 sound sources collected, the band was limited to 3 types and expressed as Sound Color Marker. Table 1 shows the frequency classification method in which the band is limited and expressed in sound color.

Table (1) Frequency Classification of Sound Color Marker according to Bandwidth-limited

Color	Bandwidth-limited		
	~ 4KHz	~ 20KHz	4KHz ~ 20KHz
Red	0 ~ 20	0 ~ 100	4000 ~ 5000
Orange	20 ~ 50	100 ~ 250	5000 ~ 6000
Yellow	50 ~ 125	250 ~ 600	6000 ~ 8000
Green	125 ~ 300	600 ~ 1500	8000 ~ 10000
Blue	300 ~ 800	1500 ~ 3500	10000 ~ 13000
Indigo	800 ~ 2000	3500 ~ 8500	13000 ~ 16000
Purple	2000 ~ 4000	8500 ~ 20000	16000 ~ 20000

Using the Sound Color Marker thus processed, similarities were compared for all sound sources in order to check whether they are the same person or not. Similarity between two sound sources expressed by Sound Color Marker is calculated by the following equation (1). In this way, the similarity comparison result is shown in Fig. 2 when Bandwidth-limited is 4Kz or less.

$$\text{Similarity between sound sources (\%)} = 1 - \sum_{k=1}^{7} |\chi(k) - y(k)| \quad (1)$$

$\chi(k)$, $y(k)$: The ratio of energy in the k bands of the two sources to be compared(%)

k: Order of frequency band of sound color marker

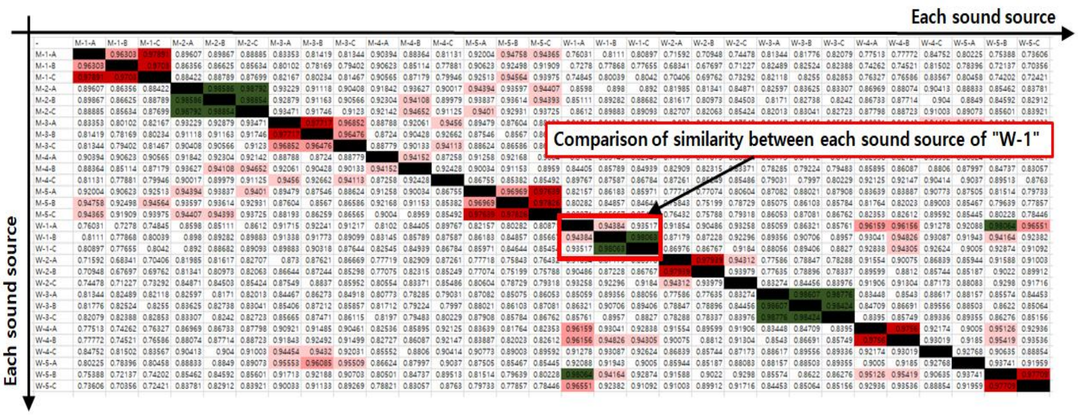

Fig 2. In the Case of Bandwidth-limited below 4KHz, Similarity between Sound Sources

Fig.2 shows the similarity of each of 30 sound sources. When the similarity is more than 94%, it is displayed in color. As the similarity is higher, it is displayed in dark color or green, and the same sound source is expressed in black color. Fig. 2 shows that the similarity is very high when the different sound sources of the same person are compared. However, it can be seen that the similarity between the sound sources that are not the same person is very low. The result of Fig. 2 shows that the probability of identification with the same person is 87% when comparing the sound sources of the same person when the threshold of the similarity is 93%. Also, when comparing the sound sources of other people, the probability of identification with others is 91%. If the threshold is increased, the probability of identifying the same person is lower, and the probability of identifying another person is higher. Fig. 3 and Fig. 4 show the similarity comparison of In the case of bandwidth-limited below 20KHz and In the case of bandwidth-limited 4 ~ 20KHz.

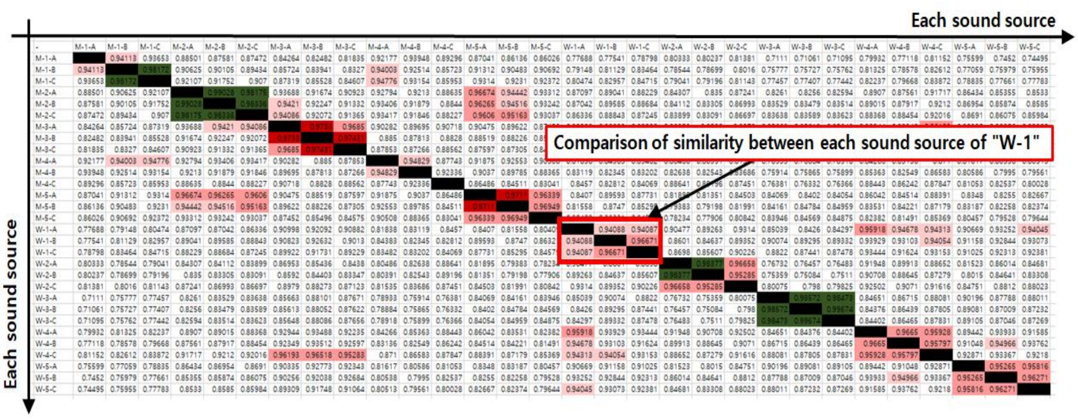

Fig 3. In the Case of Bandwidth-limited below 20KHz, Similarity between Sound Sources

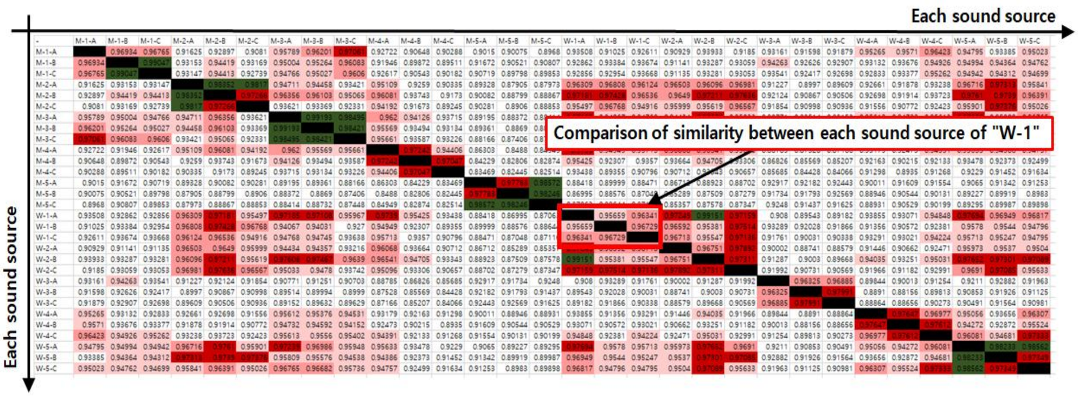

Fig 4. In the Case of Bandwidth-limited 4~20KHz, Similarity between Sound Sources

In Fig.2 ~ Fig.4, Fig.2, which is limited to 4KHz band or less, and Fig.3, which is limited to 20KHz band or less, when the threshold of similarity is 94%, It can be seen that it is somewhat identified with the same person or others. It can be seen that it is somewhat identified with the same person or others. However, in Fig. 4 where the band is limited in the interval of 4 ~ 20KHz, when the threshold of the similarity is set to 94%, identification is successful for all the same persons, but identification for others is very often failed. Based on the similarity analysis results of Fig. 2 ~ Fig. 4, the identification rate of each other is shown in Table 2.

Table (2) Identification Rate Calculation Result according to Bandwidth-limited of Voice Signal

Bandwidth -limited	Identification target	Threshold for Similarity					
		93%	94%	95%	96%	97%	98%
~ 4KHz	Identification rate for the same person	87%	73%	63%	63%	50%	23%
	Identification rate for other people	91%	94%	98%	99%	100%	100%
~ 20KHz	Identification rate for the same person	93%	90%	77%	60%	37%	27%
	Identification rate for other people	90%	95%	98%	99%	100%	100%
4 ~ 20KHz	Identification rate for the same person	100%	100%	97%	93%	67%	33%
	Identification rate for other people	53%	64%	74%	85%	94%	100%

Table 2 shows that when the threshold for similarity is 93%, when the bandwidth is limited below 4 KHz, the identification rate for the same person is 87% and the identification rate for other people is 91%. Also, when Bandwidth-limited is below 20KHz, identification rate for the same person is 93% and identification rate for other people is 90%. However, when Bandwidth-limited is applied at 4 ~ 20KHz, identification rate for other people can be expected to be more than 90% if the threshold for similarity is set to 96% or more. In the case of identification rate for other people 90%, the threshold value at Bandwidth-limited at 4KHz was 92.9% and the identification rate for the same person was 87%. In addition, the threshold value at Bandwidth-limited at 20 KHz was 93% and the identification rate for the same person was 93%. In addition, the threshold value for Bandwidth-limited at 4 ~ 20KHz was 96.5% and the identification rate for the same person was 87%. In this case, the same identification rate was exhibited when Bandwidth-limited at 4KHz and Bandwidth-limited at 4 ~ 20KHz.

IV. CONCLUSION

Human voice is a very good biometric such as security, convenience, and economy. Since human voice contains a wide variety of information, information that can distinguish individuals according to how the bands are divided is also different. In this study, we investigated identification rate difference according to Bandwidth-limited using Sound Color Marker which expresses voice signal into 7 colors. As a result of experimenting with three types of band limitation for voice signals, When Bandwidth-limited is below 20KHz and the threshold of similarity is 93%, identification rate for the same person is 93% and identification rate for other people is 90%. And the identification rate was the best. In the case of identification rate for other people 90%, identification rate for the same person is 87% when 4KHz and 4 ~ 20KHz are bandwidth-limited. However, when limiting to 4KHz, the threshold of similarity is 92.9% , And the threshold of similarity was 96.5% in case of limiting to 4 ~ 20KHz. Therefore, it can be said that the discrimination power is somewhat lowered when it is limited to 4 ~ 20KHz.

As a result of this experiment, it was confirmed that the identification rate is the best when identifying the same person or others with the Sound Color Marker when the band is limited to 20KHz or less and many components of the voice are included. There were some differences in the discrimination power depending on the method of limiting the band. However, even though the voice signal is simplified by using the Sound Color Marker, it is confirmed that all of the Bandwidth-limited methods can be identified by only the comparison of the Sound Color Marker.

The Sound Color Marker simplifies the voice signal and expresses the difference in personality information according to Bandwidth-limited. It is expected that a sound biometric authentication method suitable for various cases will be developed using Sound Color Marker.

REFERENCES

[1] Jae-Hun Song and In-Seok Kim, "A Study on the Utilization of Biometric Authentication for Digital Signature in Electronic Financial Transactions : Technological and Legal Aspect, The Journal of Society for e-Business Studies", Vol.21, No.4, pp.41-53, November 2016.

[2] S. H. Kim, Y. S. Joo, D. S. Chi, "FinTech Era: Needs for the innovation of user authentication technologies", Communications of the Korean Institute of Information Scientists and Engineers, Vol.33, No.5, pp.17-22, May 2015.

[3] S. T. Lee. Principles and Application of Sound, Cheong Moon Gak, Korea (2004).

[4] H. W. Park, S. H. Jee and M. J. Bae, "Study on The Confidence-Parameter Estimation Through Speech Signal", Journal of Multimedia Services Convergent with Art, Humanities, and Sociology, Vol.6, No.7, pp.101-108, 2016.

[5] M. J. Bae and S. H. Lee. Digital Speech Analysis, DongYoung, Korea (1998).

[6] S. G. Bae, B. M. Lim and M. J. Bae, "A study on Kim Jung-un's Obesity through Speech Signal Correlation Analysis", Journal of Multimedia Services Convergent with Art, Humanities, and Sociology, Vol.6, No.7, pp.109-116, 2016.

[7] Bong-Young Kim, and Myung-Jin Bae, "On an Analysis of the Korean Presidential Candidate, Ahn Cheol-soo's Speech." Advanced Science Letters 24.1, pp.754-756, 2018.

[8] Eun-Young Yi and Myung-Jin Bae, "A Study on A Color Expression of Sound", Conference Proceedings of Convergence Research Letter, Convergent with Art, Humanities, and Sociology, Vol.2, No.4, pp.1671-1674, October 2016.

[9] Bong-Young Kim, Eun-Young Yi, and Myung-Jin Bae, "A Study on Sound A Color about Distinguishing Voices Characteristic", Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology, Vol.8, No.2, pp.13-21, February 2018.