Biometric Bit String Generation for Smart-Phones Using Voice Data

Aki Harada, Yasushi Yamazaki* and Tetsushi Ohki†
* The University of Kitakyushu, Fukuoka, Japan
E-mail: y7mca013@eng.kitakyu-u.ac.jp, y-yamazaki@kitakyu-u.ac.jp Tel/Fax: +81-93-695-3259
† Shizuoka University, Shizuoka, Japan
E-mail: ohki@inf.shizuoka.ac.jp Tel/Fax: +81-53-478-1471

***Abstract*—Because smart-phones are in wide use, biometric authentication has attracted substantial attention as a user authentication technology for them. Voice is one of the most promising biometric modalities for user authentication on smart-phones because a user interface involving a microphone is commonly used on such devices and because biometric-specific sensors are not necessary. However, biometric authentication requires template-protection technologies to enhance the security of the biometric template that stores a user's biometric information. Therefore, to design a more secure and convenient template-protection method for smart-phones, we propose a method to generate biometric bit strings from text-independent voice data. We demonstrated its effectiveness through simulation experiments.**

I. Introduction

With the rapid spread of smart-phones, user authentication for privacy protection is becoming increasingly important. A smart-phone is equipped with various sensors, such as microphones, cameras, and touch panels, many of which enable biometric information to be obtained. Therefore, biometric authentication has attracted much attention as a user authentication technology for smart-phones [1]. One of the most promising biometric modalities for user authentication on smart-phones is voice because a user interface involving a microphone is commonly used on such devices and because biometric-specific sensors (e.g., fingerprint sensors) are not necessary.

However, in biometric authentication, a user's information is difficult to replace when a template that stores it is leaked because such information is limited. Therefore, various methods have been proposed for improving the security of templates by appropriately managing them [2].

In biometric authentication using template-protection technologies, methods allowing some errors in biometric information due to the application of an error-correcting code are currently used [3], [4]. With these methods, the biometric information must be calculated and represented on a finite field. However, individuality is lost in many current methods when the biometric information is converted into values on the finite field (hereinafter, called biometric bit strings), which may degrade authentication accuracy. Therefore, various methods for generating biometric bit strings have been proposed for converting the information into bit strings while maintaining individuality. These methods require different techniques depending on the modality or expression of biometric features. For example, an iris-based technique [5], fingerprint-based techniques [6], [7], and handwritten signature-based techniques [8], [9] have been proposed.

Voice-based techniques have also been proposed [10], [11], [12], [13], and they are the main focus of attention in this paper. However, these methods have the following problems in that using a smart-phone was not taken into account. First, voice samples used for evaluation are not always collected using a smart-phone. In these studies, experiments were carried out on a certain scale of speech database; however, details involving the method of collecting voice samples, such as the microphone used and the existence of ambient noises, were not described. We assume that voice samples collected using a microphone on a smart-phone and ones having ambient noises are indispensable for a performance evaluation. Another problem is that voice samples used for evaluation comprise a restriction on the kinds of words. Although many of these studies assume a text-independent approach, experiments were often conducted on a digit-corpus database. We assume that a text-independent approach based on free speech data collected on a smart-phone when a user is on the phone and also off the phone is more suitable for the performance evaluation. Finally, the performance of generated biometric bit strings was not always evaluated precisely. In these studies, the performance of the proposals was suitably evaluated from the viewpoint of template-protection technologies; however, the properties of biometric bit strings, which affect the performance of template-protection, were not discussed sufficiently. We assume that they are applicable to various template protection technologies if stable biometric bit strings can be generated while evaluating their properties.

Therefore, to design more secure and convenient template-protection methods for smart-phones, we proposed a method to generate biometric bit strings from text-independent voice data. We herein demonstrate its effectiveness through some simulation experiments.

The remainder of this paper is organized as follows. In Section II, our method for generating biometric bit strings is described. Next, the experimental results are presented in Section III. Finally, the conclusions are stated in Section IV.

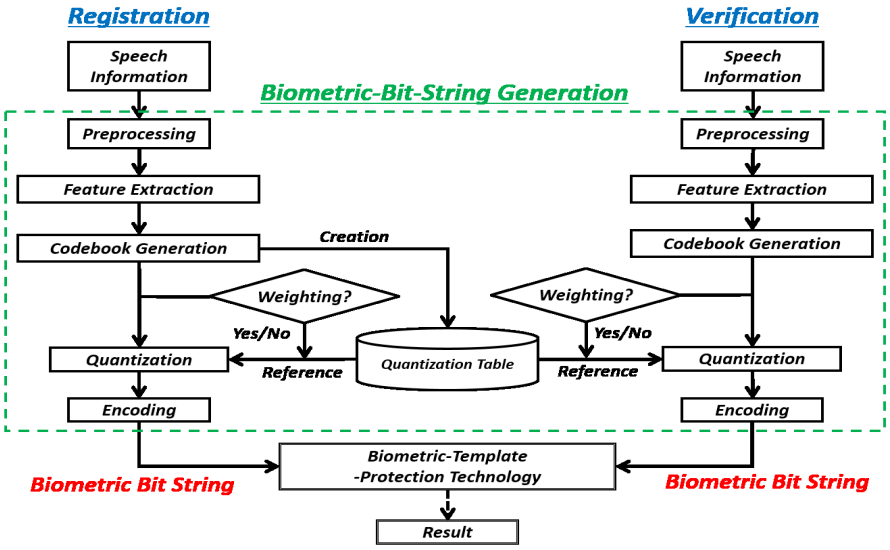


Fig. 1. Overview of biometric-bit-string generation

II. Biometric-bit-string Generation

As described in the previous section, we propose a method to generate biometric bit strings from voice data. This method generates stable biometric bit strings with arbitrary lengths from a view of their application to a general biometric template protection scheme [14] using speech information acquired in various real usage environments. Figure 1 shows an overview of the method. The main part of the method consists of preprocessing, feature extraction, codebook generation, quantization, and encoding. In the following subsections, we describe how biometric-bit-strings were generated from voice data on a smart-phone.

A. Preprocessing

We removed noise and long sections of silence in the voice data obtained from a microphone on a smart-phone. We performed them using noise reduction and truncating silence in Audacity [15], which is free, open-source software. Table I show the threshold values that we used in accordance with each effect function.

TABLE I
Preprocessing parameters

Noise reduction	Noise reduction		12 dB
	Sensitivity		6.00
	Frequency smoothing		3 bands
Truncating silence	Detect silence	Level	-45 dB
		Duration	0.01 sec.
	Action	Truncate to	0.001 sec.

B. Feature Extraction

We used the Mel frequency cepstral coefficients (MFCC) technique to extract features from the voice data. Starting with an energy-based voice activity detection, MFCC features were extracted from the speech signal using a window function of 20-30 ms. We obtained (N =) 12 coefficients and the log-energy value for each frame and the first and second-order derivative, i.e., we obtained a feature vector with (Z =) 39 components per frame ($N = \frac{Z}{3} - 1$) in total.

C. Codebook Generation

A codebook was generated using the LBG algorithm, which is the typical algorithm of vector quantization (VQ) [16], for the feature vectors extracted in the previous section. With the proposed method, we set the codebook level M to 32 on the basis of the preliminary experiments. The level was set based on the fact that no further improvement in accuracy was observed. Figure 2 shows the components of the generated feature vector.

$$\boldsymbol{x}_i = (\underbrace{x_{i,1}, \cdots, x_{i,N}}_{\text{MFCC}}, \overbrace{x_{i,N+1}}^{\text{Power}}, \underbrace{x_{i,N+2}, \cdots, x_{i,2N+1}}_{\Delta\text{MFCC}}, \overbrace{x_{i,2N+2}}^{\Delta\text{Power}}, \underbrace{x_{i,2N+3}, \cdots, x_{i,3N+2}}_{\Delta\Delta\text{MFCC}}, \overbrace{x_{i,Z}}^{\Delta\Delta\text{Power}}) \quad (1 \le i \le M)$$

Fig. 2. Components of feature vector

D. Weighting In Consideration Of Individuality

Individuality may not be included equally in each dimension of a codebook. Therefore, we assigned a bigger bit length for the dimension in which individuality is greatly included. First, we calculated the variance $\sigma_j^2 (1 \le j \le Z)$ in each dimension of the codebook of each voice datum as part of a concrete technique. The smaller the variance, the more stable the corresponding features obtained from the user. Table II and

Figure 3 show an overview of weighting in consideration of individuality.

TABLE II
VARIANCE FOR EACH CODEBOOK

Dimension	1	2	···	N	N+1	···
$\boldsymbol{x_1}$	$x_{1,1}$	$x_{1,2}$	···	$x_{1,N}$	$x_{1,N+1}$	···
$\boldsymbol{x_2}$	$x_{2,1}$	$x_{2,2}$	···	$x_{2,N}$	$x_{2,N+1}$	···
⋮			⋮			
$\boldsymbol{x_M}$	$x_{M,1}$	$x_{M,2}$	···	$x_{M,N}$	$x_{M,N+1}$	···
Variance	σ_1^2	σ_2^2	···	σ_N^2	σ_{N+1}^2	···

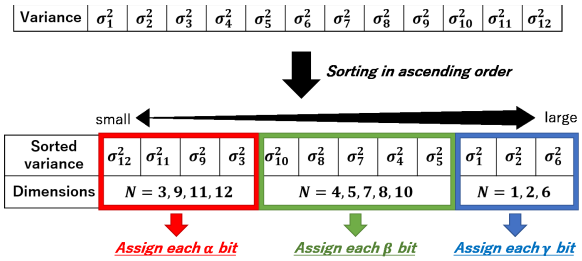


Fig. 3. Example of bit allocation (ex. MFCC)

We sorted these values in ascending order for each feature vector, such as MFCC ($1 \leq j \leq N$), ΔMFCC ($N+2 \leq j \leq 2N+1$), and ΔΔMFCC ($2N+3 \leq j \leq 3N+2$). The preliminary experiments showed a large difference in variance between the 4th and 5th dimensions of the feature vector, whose components were sorted in ascending order of variance. Also, a large difference occurred between the 9th and 10th dimensions of the feature vector. Therefore, we divided its components into three groups, where the variances were close to each other within the same group, and assigned α, β, and γ bit ($\alpha > \beta > \gamma$) to each group. We set $\alpha = 4$, $\beta = 3$, and $\gamma = 2$ in this proposal by assuming that it will be implemented on smart-phones in the future.

E. Quantization

To reduce the fluctuation in the extracted features, we quantized each element of each representative vector by referring to the preset quantization table shown in Table III. Parameter $a_k (1 \leq k < Q)$ was determined for each feature in advance on the basis of preliminary experiments so that the observation frequency of the feature value was equal over each of the quantization steps. Using the proposed method without weighting in consideration of individuality, we set the quantization level to 16 on the basis of the preliminary experiments. However, when weighting was done in consideration of individuality, we set the quantization level to 16, 8, and 4 on the basis of those experiments. The level was determined on the basis of the stability of the feature values in the experiments and the fact that the bit length generally used as an encryption key in the field of cryptography was often 2^n.

TABLE III
QUANTIZATION TABLE

x_{ij}	$S\ [x_{ij}]$
$x_{ij} < a_1$	0
$a_1 \leq x_{ij} < a_2$	1
⋮	⋮
$a_{Q-1} \leq x_{ij}$	$Q-1$

F. Encoding

We encoded $S\ [x_{ij}]$ into bit strings consisting of "0" and "1." In the proposed method, we expressed the quantized value in Gray code. Because Gray code always has a Hamming distance of 1 between adjacent bit strings, we can represent features using the same bit or 1-bit difference if their values are close.

Four bits were obtained from each feature because our quantization level was 16 in this process. In this paper, we define the concatenated bit string over all features as a biometric bit string.

III. Simulation Experiments

To evaluate the method's reliability, we conducted two simulation experiments. In the first, we evaluated the performance of the generated biometric bit strings from the viewpoint of verification accuracy. In the second, we evaluated the statistical features of the generated biometric bit strings. Table IV lists the experimental conditions, and Table V lists the types of noises used in the experiments.

TABLE IV
EXPERIMENTAL CONDITIONS

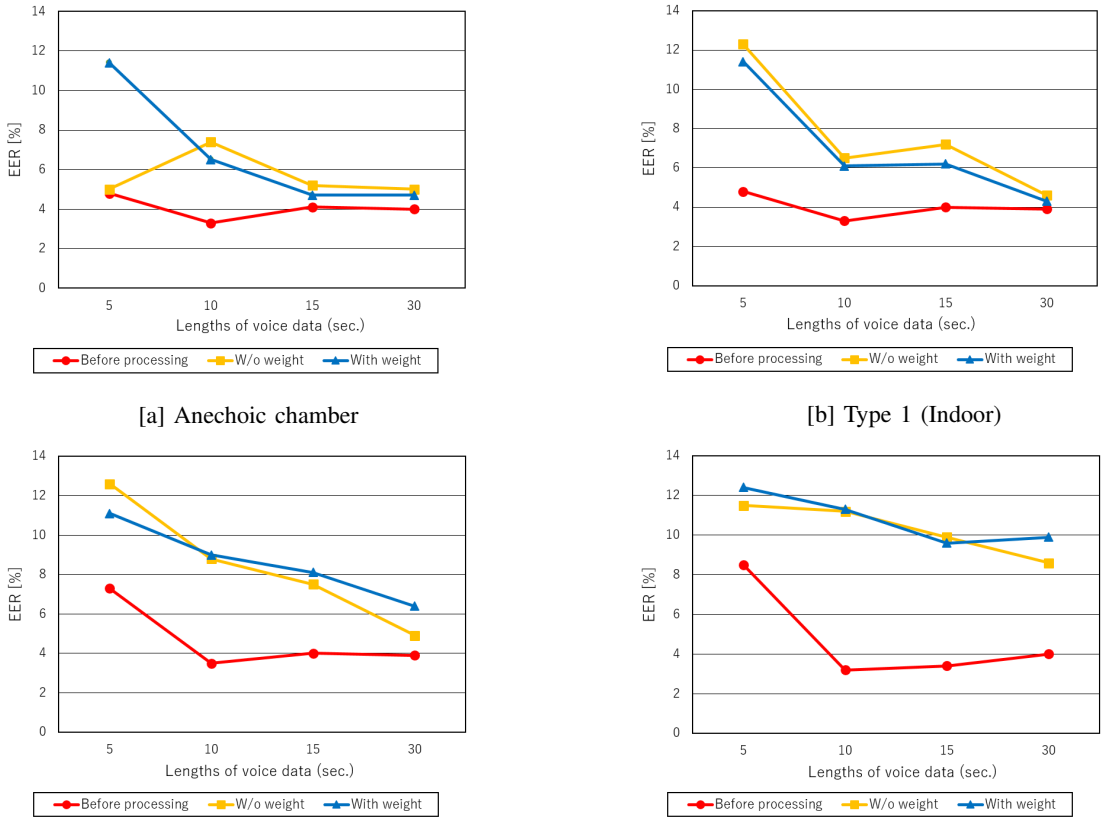
Number of speakers	10 (9 male and 1 female)
Specification	Japanese free speech
Registration data	60 sec. speech × 6 times
Verification data	θ sec. speech × 6 times (θ = 5, 10, 15, 30)
Device	ASUS Zenfone 2 Laser
Sensor	Built-in microphone (44.1 kHz, 16 bit)
State of device	Assuming a call
Software used	• Speech Signal Processing Toolkit (SPTK) version 3.9 • Audacity version 2.2.2

TABLE V
TYPES OF NOISES

Type	Condition		Noise level (dB)		
			Min.	Max.	Avg.
1	Indoor	W/o air conditioning noise	33.2	35.2	33.9
2		With air conditioning noise	45.5	47.5	46.4
3	Outdoor	W/o traffic noise	45.4	56.8	48.6

A. Accuracy of Individuality of Biometric Bit String

In this experiment, we evaluated the performance of the generated biometric bit strings from the viewpoint of verification accuracy. We calculated the equal error rate (EER), the value where the false match rate (FMR) and false non-match rate (FNMR) are equal. Figure 4 lists the results of the experiment under different types of noises and different lengths of data. In Figure 4, the column of "before processing" denotes the results



[a] Anechoic chamber

[b] Type 1 (Indoor)

[c] Type 2 (Indoor with air conditioning noise)

[d] Type 3 (Outdoor)

Fig. 4. EER for each noise condition

when we evaluated the performance in terms of the feature-vector level before converting to the bit strings. However, the columns of "without weight" and "with weight" denote the results when we evaluated the performance in terms of bit-string level. As shown in Figure 4, the verification accuracy was degraded when the data were very short (5 sec. for example). A comparison of the feature-vector-based results with bit-string-based ones shows that the verification accuracy was not degraded much by quantization even in the form of bit strings when the data were sufficiently long; however, it depended on the noise type (see Type 3 in the same table). Moreover, the weights tended to be effective when the noise level was low and when the data were long. A comparison of the aforementioned results with those described in the related work in Sect.1 shows that our method is not always superior to the previous work in terms of EER; however, it has advantages over them in terms of possibility of using free speech data considering various smart-phone-specific user services in the near future.

B. Statistical Features of Biometric Bit String

In this experiment, we evaluated the statistical features of the generated biometric bit strings from the viewpoint of the potential for using private keys (cryptographic keys) for template-protection applications.

1) Randomness: We evaluated the occurrence rate of "0" of a bit string for each user's data. It's average was 50.0% $\pm$ 2.7%, and the standard deviation was 1.5 $\pm$ 0.3. Considering that a random string had a theoretical average rate of 50% and a standard deviation of 5.1 and that the bit string was as long as 4992 bits (w/o weight) or 3936 bits (with weight), we found that the generated biometric bit strings also retained sufficient randomness with the occurrence frequency of "0" and "1."

2) Correlation: We evaluated the auto-correlation and cross-correlation of biometric bit strings. The average and variance of correlation values of the bit strings with phase shift generated from the same users were about 2.0×10^{-3} and about 4.0×10^{-4}, demonstrating the effectiveness of generated biometric bit strings in the proposed method. Moreover, the average and variance of correlation values of the bit strings with phase shift generated from the different users were about 2.0×10^{-3} and about 4.0×10^{-4}, demonstrating the biometric bit strings clearly differ between users.

IV. Conclusions

To design a more secure and convenient template-protection method for smart-phones, we proposed a method to generate biometric bit strings from voice data collected on smart-phones. The simulation results suggest the proposed method is effective.

The uniqueness and entropy of the generated biometric bit strings should be evaluated to apply them to certain template-protection applications in the future. In addition, because the proposed method utilized a VQ-based speaker model for simplicity, more sophisticated speaker models such as Gaussian mixture models and universal background models need to be considered. Such models are expected to be more robust against noisy environments. Moreover, from the viewpoint of template-protection technology suitable for smart-phones with limited computational complexity and storage capacity, the reliability of our method needs to be evaluated in more real environments.

Acknowledgments

Part of this work was supported by JSPS KAKENHI Grant Number JP16K00190.

References

[1] P.A. Tresadern, C. McCool, N. Poh, P. Matejka, A. Hadid, C. Levy, T.F. Cootes, and S. Marcel, "Mobile Biometrics: Combined Face and Voice Verification for a Mobile Platform," IEEE Pervasive Computing, 12, 1, pp. 79-87, 2013.

[2] A.K. Jain, K. Nandakumar, and A. Nagar, "Biometric Template Security," in *EURASIP Journal on Advances in Signal Processing*, 1, pp. 7-13, 2008.

[3] A. Juels and M. Wattenberg, "A fuzzy commitment scheme," in *Proc. Sixth ACM Conf. Computer and Communications Security (CCS '99)*, pp. 28-36, 1999.

[4] A. Juels and M. Sudan, "A fuzzy vault scheme," in *Proc. IEEE Int. Symp. Information Theory (ISIT 2002)*, p. 408, 2002.

[5] C. Rathgeb and A. Uhl, "Adaptive fuzzy commitment scheme based on iris-code error analysis," in *Proc. 2nd EUVIP'10*, pp. 41-44, 2010.

[6] K. Nandakumar, A. Nagar, and A.K. Jain, "Hardening Fingerprint Fuzzy Vault Using Password," in *Proc. of ICB2007*, LNCS, 4642, pp. 927-937, 2007.

[7] Z. Jin, T.S. Ong, C. Tee, and A.B.J. Teoh, "Generating revocable fingerprint template using polar grid based 3-tuple quantization technique," in *Proc. 54th IEEE Int. Midwest Symp. Circuits and Systems (MWSCAS 2011)*, pp. 1-4, 2011.

[8] H. Feng and C.C. Wah, "Private key generation from on-line handwritten signature," Information Management and Computer Security, pp. 159-164, 2002.

[9] C. Vielhauer, R. Steinmetz, and A. Mayerhőfer, "Biometric Hash based on Statistical Feature of Online Signature," in *Proc. 16th Int. Conf. Pattern Recognition*, 1, pp. 123-126, 2002.

[10] S. Billeb, C. Rathgeb, H. Reininger, K. Kasper, and C. Busch, "Biometric template protection for speaker recognition based on universal background models," IET Biometrics, Vol. 4, Iss. 2, pp. 116-126, 2015.

[11] M. Paulini, C. Rathgeb, A. Nautsch, H. Reichau, H. Reininger, C. Busch, "Multi-Bit Allocation: Preparing Voice Biometrics for Template Protection," in *Proc. Odyssey 2016*, pp. 291-296, 2016.

[12] K. Inthavisas and D. Lopresti, "Secure speech biometric templates for user authentication," IET Biometrics, Vol. 1, no. 1, 2012.

[13] R. C. Johnson and T. E. Boult, "With vaulted voice verification my voice is my key," in *Proc. Int'l Conf. on Technologies for Homeland Security (HST' 13)*, 2013.

[14] ISO/IEC 24745, "Information technology - Security techniques - Biometric information protection," ISO/IEC, 2011.

[15] [Online]. Available: https://manual.audacityteam.org/index.html

[16] Y. Linde, A. Buzo, and R.M. Gray, "An algorithm for vector quantizer design," IEEE Trans. Commun., COM–28, 1, pp. 84-95, 1980.