ELectronics EXpress

IEICE Electronics Express, Vol.2, No.7, 226–231

Multi-space random mapping for speaker identification

Chong Tze Yuang [a)], **Andrew Teoh, David Ngo, and Michael Goh**

Centre of Biometrics & Bioinformatics,

Faculty of Information Science & Technology,

Multimedia University,

Jalan Ayer Keroh Lama, 75450 Melaka, Malaysia

a) *tychong@mmu.edu.my*

Abstract: This paper presents a work of utilizing multi-space random mapping (MRM) to formulate a dual-factor identification system, which combines speaker biometric and personal token. Our work has shown that MRM-system exhibits stronger discriminative ability when comparing test features to its counterfeit templates, which lied in other different random subspaces. This advantage thus contributes to better F-ratio and greater accuracy recognition.

Keywords: Speaker identification, multi-space random mapping, dual-factor authentication

Classification: Science and engineering for electronics

References

[1] W. B. Johnson and J. Lindenstrauss, "Extension of Lipschitz mapping into a Hilbert Space," *Conf. in Modern Analysis and Probability*, Amer. Math. Soc., Providence, R.I., pp. 189–206, 1984.

[2] J. Makhoul, "Linear Prediction: A tutorial review," *Proc. IEEE*, vol. 63, pp. 561–580, 1975.

[3] F. K. Soong, A. E. Rosenberg, L. R. Rabiner, and B.-H. Juang, "A vector quantization approach to speaker recognition," *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 387–390, 1985.

[4] A. Higgins, L. Bhaler, and J. Porter, "Voice identification using nearest neighbor distance measure," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 375–378, 1993.

[5] A. Higgins, "YOHO speaker verification," *Speech Research Symp.*, 1990.

[6] A. B. J. Teoh, D. C. L. Ngo, and A. Goh, "Quantized multispace random for two-factor identity authentication," *IEEE Trans. Pattern Anal. Machine Intell.*, under review.

[7] M. Savvides, B. V. K. V. Kumar, and P. K. Khosla, "Cancelable biometric filters for face recognition," *Proc. 2004 ICPR*, vol. 3, pp. 922–925, 2004.

1 Introduction

Traditional token-based authentication system, e.g., smart card etc. has the major drawback of being easily fooled by the stolen token. Intruder who is

DOI: 10.1587/elex.2.226
Received March 02, 2005
Accepted March 23, 2005
Published April 10, 2005

holding a stolen token will trespass the security easily. Speaker recognition emerged as the more reliable authentication system based on the assumption that human speaking behavior is unique from others thus can be utilized as the biometric feature for authentication. However, the variation of the speaking manner of human is the natural disadvantage of speaker-biometric making it the least accurate biometrics compare to other static-signal biometrics such as fingerprint etc.

By combining the personal token and the speaker biometric, attacker can no more breaks through the security simply by presenting an stolen token. Since the personal token is a very unique factor, it can be hashed into the speaker biometrics, in some manner, to make the speaker feature more distinctive. Similar works can be seen in [6, 7] where the personal token is introduced to a trademark biometrics system to cure some major defect in the current security system. This paper records the work of combining the token-based authentication with the speaker recognition to gain benefit from both sides and to alleviate the drawback of them. Multi-space random mapping (MRM) is used to hash the token information to the speaker feature.

2 Multi-space random mapping

Multi-space random mapping (MRM) composes two stages: (a) feature extraction and (b) random projection. The feature extracted from speech raw signal is mapped to a client specified random subspace. The mapping is determined by the tokenized pseudo random numbers (PRN), which are unique from a speaker to another speaker. The random projection can be expressed as follows [1]:

$$v = \kappa R \omega. \tag{1}$$

Vector ω is the original p-order feature and R is the $p \times p$ row-wise orthonormal random matrix built from the PRN. The value of κ is unity since the feature dimension is remained in this projection.

During enrollment, p^2 numbers of standard normal distributed PRN are generated and are assigned to the new-registered speaker. The PRN are arranged into a $p \times p$ matrix to be orthonormalized using singular vector decomposition (SVD) algorithm to construct the transformation matrix, R. Each row of R essentially is the axis of the speaker-specified random subspace. Features extracted from the training speech will be mapped to the random subspace and speaker template will be produced from those random-mapped features. Thus, template of different speaker is spanned in different random subspace. Identification is performed by mapping the test features to each of the registered subspace to match to the respective template. The arriver will be identified as the speaker whose template yields the closest distance to the test features.

Since the transformation is orthonormal, the distance between any two points that are projected to the same subspace is identical to the original distance before mapping, e.g.,

$$\|T_1 a - T_1 b\|^2 = \|a - b\|^2 . \tag{2}$$

DOI: 10.1587/elex.2.226
Received March 02, 2005
Accepted March 23, 2005
Published April 10, 2005

Therefore the similarity between the test features and the genuine template is not altered after MRM is applied to the speaker authentication system thus preserves the intra-speaker distance.

However, by mapping two points into two different subspaces, we have found that it is almost certain that the distance between the two points will be stretched, e.g.,

$$P\left(\|T_1 a - T_2 b\|^2 > \|a - b\|^2\right) \simeq 1. \tag{3}$$

While the searching for the mathematical evidence to verify Eq. (3) is still being carried out, Eq. (3) has been proven empirically and the detail of the experiment will be discussed in the next section. Under MRM scheme, the test features become more discriminative to the counterfeit templates thus raises the inter-speaker distance overall. Therefore, the MRM-speaker authentication shows better F-ratio, e.g., ratio of intra-speaker to inter-speaker distance, as compare to the non-MRM version.

3 Speaker identification system

The block diagram shown in Fig. 1 depicts the speaker identification system in this work. MRM scheme is built on the vector quantization (VQ) speaker recognition framework [3]. Collected speech waveform is blocked into 240-sample frame with 160 overlapped with adjacent frames. Each frame is categorized into speech and non-speech frame based on the energy profile. Speech frames are selected to extract the linear predictive (LP) cepstrum [2].

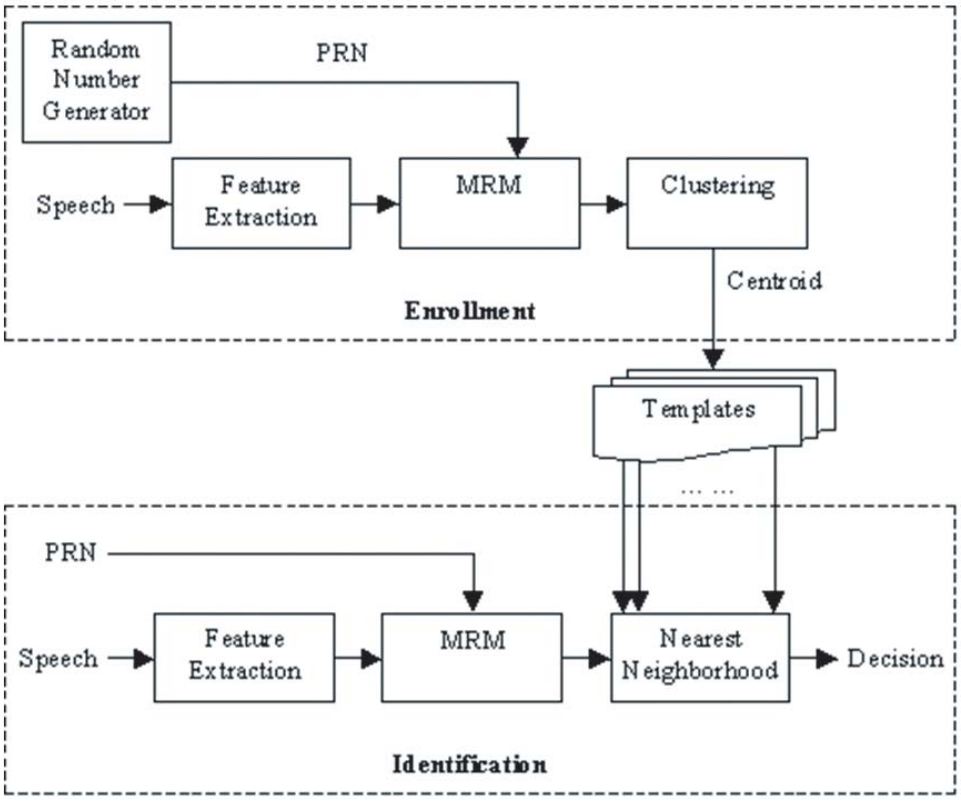

Fig. 1. Block diagram of speaker identification system.

DOI: 10.1587/elex.2.226
Received March 02, 2005
Accepted March 23, 2005
Published April 10, 2005

During enrollment, a new PRN sequence is generated to form the random transformation to map the features extracted from the training speech signal. The feature vectors in the new subspace are clustered into 16 clusters using modified K-means (MKM) algorithm [3]. Each cluster contributes one centroid to be stored as the speaker's template.

In the identification session, test speech waveform and the PRN sequence are inputted to the identifier. Feature extracted from the test speech is mapped to the random subspace that is described by the PRN. Nearest-neighborhood matching is performed to compare the projected test feature to each registered templates [4] in each random subspace.

4 Database

The experiment was conducted on YOHO speaker verification corpus [5]. The speech tokens from the first enrollment session are used to generate the template from all 138 speakers. All speech tokens from testing session are used to evaluate the system. Thus each speaker template is generated from 24 tokens and there are total 5,520 (40 tokens $\times$ 138 speakers) trails of identification testing.

5 Experiments and Discussion

The fact that there is such a great possibility that the distance between two points will be stretched resulted from MRM, as stated in Eq. (3), has been proven through experiments. The experiment has been carried out by running 50,000 trials to collect the amount of distance that is extended between two points resulted from MRM. Two points, a_i and b_i, are picked randomly in every trial and to be mapped into two different random subspaces. The random mapping is described by the orthonormal random matrixes, R and T, which are refreshed in every trial. Table I tabulates the mean and the standard deviation of the original distance, $d_i = \|a_i - b_i\|^2$, projected distance, $m_i = \|R_i a_i - T_i b_i\|^2$, and the amount of extension of the distance resulted from MRM, e.g.,

$$\begin{aligned} e_i &= m_i - d_i \\ &= \|R_i a_i - T_i b_i\|^2 - \|a_i - b_i\|^2 . \end{aligned} \tag{4}$$

Experiment is carried out in the dimension of 10, 15, 20, 30, and 50 orders.

Table I. Extension of the distance in MRM

Order	Original		Projected		Extension		$n(e>0)$	$P(e>0)$
	$\overline{d}$	σ_d	$\overline{m}$	σ_m	$\overline{e}$	σ_e		
10	1.669	0.625	6.668	2.497	4.999	2.527	49616	0.9923
15	2.503	0.752	9.978	3.060	7.475	3.096	49930	0.9986
20	3.333	0.884	13.306	3.518	9.979	3.565	49990	0.9998
30	4.993	1.079	19.987	4.316	14.993	4.382	49999	1.0000
50	8.330	1.386	33.324	5.565	24.994	5.643	50000	1.0000

DOI: 10.1587/elex.2.226
Received March 02, 2005
Accepted March 23, 2005
Published April 10, 2005

As observed, almost all trial stretches the distance, e.g. $e_i > 0$, therefore the following conclusion is drawn thus certifies Eq. (3), e.g.,

$$\begin{aligned} &P\left(e > 0\right) \simeq 1 \\ &P\left(\|Ra - Tb\|^2 > \|a - b\|^2\right) \simeq 1. \end{aligned} \tag{5}$$

As observed, when the experiment is repeated in the dimension of higher orders, more trials were experienced distance extension. At the order of 50, distance is extended in all 50,000 trials. Thus, we deduce a conclusion that it is more certain that MRM extends the distance at higher order of dimension.

In order to monitor the effect of MRM to the speaker identification system, experiments are carried out to compare the performance of the MRM-system with the non-MRM system. Without engaging MRM, the identification system follows the traditional VQ-speaker recognition framework [3]. The accuracy and the F-ratio of the system at different feature order are shown in Table II as follows.

Table II. Performance of the system with MRM and without MRM

Order	Identification rate (%)		F-ratio	
	MRM	non-MRM	MRM	non-MRM
8	100	71.0	0.171	0.585
10	100	76.4	0.163	0.575
12	100	79.8	0.156	0.587
14	100	81.6	0.151	0.602
16	100	82.6	0.148	0.615
18	100	83.8	0.146	0.630
20	100	83.3	0.144	0.650

The MRM-system scores the perfect identification rate at feature order from 8 to 20. For MRM-identification system, higher feature order yields better F-ratio. This is contrast with the manner of the F-ratio shown by the original system, where higher order produces weaker F-ratio. The phenomenon is due to the behavior of MRM where higher order MRM stretches greater distance between two points. At higher feature order, the inter-speaker matching yields better discrimination thus contributing to lower F-ratio.

6 Conclusion

In this work, MRM system has been shown to possess remarkable discriminative ability in counterfeit comparison thus contribute to better accuracy recognition. MRM has been incorporated with the classical VQ speaker identification system to form a dual-factors identification system that combines the speaker biometric and the personal token. Experiments on YOHO corpus has scored perfect identification rate.

DOI: 10.1587/elex.2.226
Received March 02, 2005
Accepted March 23, 2005
Published April 10, 2005

Acknowledgments

This project is sponsored by Multimedia University Internal Research Funding Cycle 1/2004 PR/2004/0392.

DOI: 10.1587/elex.2.226
Received March 02, 2005
Accepted March 23, 2005
Published April 10, 2005