Probabilistic Random Projections and Speaker Verification

Chong Lee Ying and Andrew Teoh Beng Jin

Faculty of Information Science and Technology (FIST), Multimedia University,
Jalan Ayer Keroh Lama, 75450 Melaka, Malaysia
lychong@mmu.edu.my
Biometrics Engineering Research Center (BERC), Yonsei University, Seoul, South Korea
bjteoh@ieee.org

Abstract. Biometrics is susceptible to non-revocable and privacy invasion problems. Multiple Random Projections (MRP) was introduced as one of the cancellable biometrics approaches in face recognition to tackle these issues. However, this technique is applicable only to 1D fixed length biometric feature vector but failed in varying size feature, such as speech biometrics. Besides, simple matching metric that used in MRP unable to offer a satisfactory verification performance. In this paper, we propose a variant of MRP, coined as Probabilistic Random Projections (PRP) in text-independent speaker verification. The PRP represents speech feature in 2D matrix format and speaker modeling is implemented through Gaussian Mixture Model. The formulation is experimented under two scenarios (legitimate and stolen token) using YOHO speech database. Besides that, desired properties such as one-way transformation and diversity are also examined.

Keywords: Speaker verification, Cancellable biometrics, Probabilistic Random Projections.

1 Introduction

Although biometrics is a powerful tool against repudiation and has been widely deployed in various security systems, the biometric characteristics are largely immutable, resulting in permanent biometric compromise. Cancellable biometrics was introduced by Ratha et al. [1] in storing a transformed version of the biometric template and provides higher privacy level by allowing multiple templates to be associated with the same biometric data. This helps to promote non-linkability of user's data stored across various databases. Basically, a good cancellable biometrics formulation must fulfill the following requirements:

1. Diversity: The same cancellable template cannot be used in two different applications.
2. Reusability: Straightforward revocation and reissue in the event of compromise.
3. Non-invertibility of template computation to prevent recovery of biometrics.
4. Performance: The cancellable biometric template should not deteriorate the recognition performance.

S.-W. Lee and S.Z. Li (Eds.): ICB 2007, LNCS 4642, pp. 445–454, 2007.

The details survey of various cancellable biometrics constructs can be found in [2]. In this paper, we focus only on the constructs which combine external factors such as tokenized random data with biometrics. Soutar et al. [3] first proposed a cancellable biometrics generated from fingerprints using optical computing techniques. The idea is to create identification codes, which are completely independent to the biometric data and can be easily modified and updated in the future. The templates consisted of correlation patterns derived from training images, which are subsequently mixed with random data to generate identification codes. However, the scheme was not explained in a satisfactory manner regarding the cryptographic security aspects of the transformations where no related results can be found. Teoh et al. [4] introduced a cancelable biometrics/key via inner product between randomized token and biometric data with quantization. This method is advantageous in comparison to that of Soutar et al since the transformation is a one-way process. Unfortunately, their formulation suffered from the scenario when the genuine token was stolen and used by the imposter to claim as the genuine user (stolen-token scenario). In this case, the recognition performance becomes poorer. This against the fourth criteria of cancellable biometrics.

Savvides et al. [5] proposed a cancellable biometrics scheme which encrypted the training images by synthesizing a correlation filter for biometric authentication. They demonstrated that convolving the training images with any random convolution kernel prior to building the biometric filter does not change the resulting correlation output peak-to-sidelobe ratios, thus preserving the authentication performance. Despite of that, the security will be jeopardized via a deterministic deconvolution with a known random kernel. Chong et al. [6] presented a work by utilizing multi-space random mapping to formulate a dual-factor speaker recognition system which combines speaker biometrics and user-specific token, but no report was given on the performance in stolen-token scenario. Note that Soutar et al. and Savvides et al. applied a set of common random numbers for all users whereas Teoh et al. and Chong et al. are using user-specific random numbers. The former aimed to conceal the biometrics data whereas the later utilized token's randomness for performance enhancement.

Recently, Teoh et al. [7] introduced the Random Multispace Quantization (RMQ), as an analytic mechanism for BioHash in face recognition, where the process is carried out by the non-invertible random projection of biometric feature and quantization. RMQ can be revoked through the pseudo-random numbers (PRN) replacement so that a new template can be generated instantly. When RMQ does not involve the quantization process, the formulation is named as Multiple Random Projections (MRP) [9]. Performance of MRP is improved through the projection from feature domain to a class-specified random subspace. Thus, the intra-class variations are preserved and enhance the inter-class variations. They inferred that the recognition performance is retained as sole biometrics performance in stolen-token scenario. This is accomplished through the choice of dissimilarity metric - normalized dot product that governs the statistic preservation transformation.

In this paper, we extend the MRP for speaker verification by using the 2D Principle Component Analysis Gaussian Mixture Model (2DPCA-GMM) [8] in speaker modeling instead of using normalized dot product. We name this method as Probabilistic Random Projections (PRP). Specifically, we remove the limitation of MRP, which is applicable only to fixed length 1D feature vectors. This is important because speech feature is varied in size due to the recording time length and it is normally represented in 2D matrix format. Beside that, we show that probabilistic treatment of MRP still survives in stolen-token scenario, subjected to certain condition. We also examine the diversity property of PRP.

This paper is organized as followed: Section 2 presents the brief introduction of Multispace Random Projections. Section 3 explains the Probabilistic Random Projections method. The experimental results and discussion are given in Section 4. The conclusion and future works are provided in Section 5.

2 Brief Reviews of Multispace Random Projections (MRP)

Multiple Random Projections (MRP) comprises two stages: (a) feature extraction and (b) random projections. In feature extraction stage, the individual's 1D feature vector, $\mathbf{x} \in \Re^{d}$, with length d is extracted. The feature vector is then projected onto a random subspace, $\mathbf{R} \in \Re^{mxd}$, $m<d$ through $\mathbf{y}= (1/\sqrt{m})\, \mathbf{Rx} \in \Re^{m}$. $\mathbf{R}$ is generated from the external sources such as pseudo-random numbers (PRN). During verification, the feature vector is mixed with genuine PRN and the resulting vector is compared with the enrolled template by using the normalized dot product.

Since MRP performs on user-specific basis; in the real world application, we should consider two different scenarios:

1. Legitimate Token: When the genuine **x** is concealed with **R**, which is generated by his specific PRN.
2. Stolen Token: in which an imposter has access genuine **R** and used by the imposter to claim as the genuine user.

We summarize the performance behavior of above two scenarios by using the genuine-imposter distribution as shown in Fig. 1.

In general, the accuracy of the biometrics system is determined by how much overlapping there is between the two distributions - genuine and imposter distributions. The larger of overlapping of two distributions, the poorer of system will be and vice versa. For the MRP in all scenarios, it shows that the statistical properties – mean and standard deviation of *genuine distribution* are preserved just like in the feature vector level (original system without random projection). On the other hand, imposter distribution is peaked at 1 and the standard deviation is equal to $1/\sqrt{m}$ in scenario 1 (Legitimate Token) by using normalized dot-product as matcher. This echoes that the clear separation of the genuine-imposter distribution can be attained, and hence near to zero error rate if m is sufficiently large as depicted in Fig. 1. However, in the second scenario, the inter-class variation shall revert to its original state in feature vector level and hence the performance is retained like before random projection is performed when $m<d$.

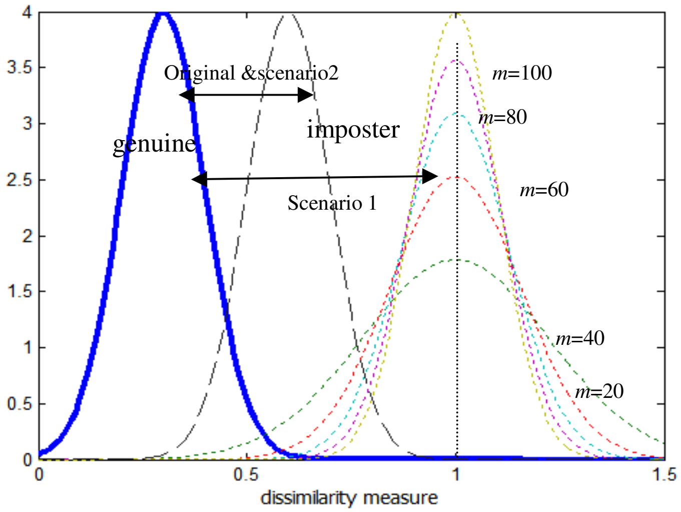


Fig. 1. Genuine-Imposter distributions for MRP

3 Probabilistic Random Projections

Probabilistic Random Projections (PRP) consists of three stages: (a) feature extraction (b) random projection (c) probabilistic modeling. The block diagram of Probabilistic Random Projections in speaker verification system is shown in Fig. 2.

The individual's feature matrix, $\mathbf{X} \in \Re^{qxr}$ is first extracted from the preprocessed framed speech signal through Linear Predictive Coding [10], where q represents the order of feature coefficient and r represents the frame number. Note that r is varied according to the recording length. Then, 2D Principal Component Analysis is used to compress **X,** hence, $\mathbf{W} \in \Re^{pxr}$ where $p \leq q$.

The 2D-PCA feature, **W** is further projected onto a random subspace as determined from an externally derived PRN, $\mathbf{R} \in \Re^{mxp}$, where $m \leq p$. The user-specific random-projected vector, $\mathbf{Y} \in \Re^{mxr}$ is obtained through the random projection process which is defined as:

$$\mathbf{Y} = \mathbf{RW} \tag{1}$$

The non-invertible property of **Y** can be assessed by referring the equation (1). **Y** can be regarded as a set of underdetermined systems of linear equations (more unknowns than equations). Therefore, it is impossible to find the exact values of all the elements in **W** by solving an underdetermined linear equation system in $\mathbf{Y} = \kappa\mathbf{RW}$ if $m<p$, based on the premise that the possible solutions are infinite.

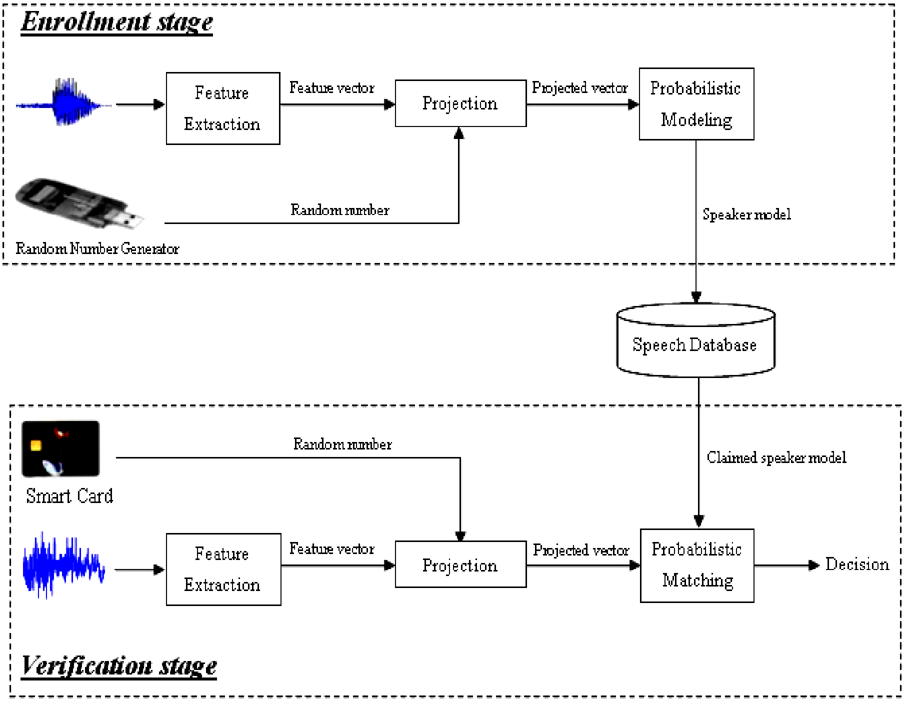


Fig. 2. Block diagram of Probabilistic Random Projections in speaker verification system

In the probabilistic modeling stage, the random-projected vector **Y** is fed into a Gaussian Mixture Model (GMM) [11] to construct a probabilistic speaker model. The speaker model produced by each speaker, λ is located in different random subspace. The Gaussian mixture density is computed by a weighted sum of *M* component densities defined as $p(\mathbf{Y}\,|\,\lambda)=\sum_{k=1}^{M} m_k b_k(\mathbf{Y})$, where *M* is the Gaussians mixture order. Mixture weights for Gaussian are represented by m_k, which satisfy the constraints $\sum_{k=1}^{M} m_k = 1$ and $b_k(\bullet)$, is a Gaussian function $\mathcal{N}(\boldsymbol{\mu}_k, \Sigma_k)$ with mean vector $\boldsymbol{\mu}_k$ and covariance, Σ_k. Each speaker is parameterized by its mixture weights, mean vectors and covariance matrices as $\lambda_k = \{m_k, \boldsymbol{\mu}_k, \Sigma_k\}$. GMM is vital to resolve the varying matrix size problem in 2D-PCA features that is not possible to be handled by simple metric used in [7].

During verification, the claimed speaker will present his speech biometrics and personal PRN. The feature extracted from the test sample will be projected to the user-specific random subspace generated from the personal PRN sequence from the claimed speaker. This new random-projected feature is then input into GMM for probabilistic matching. A likelihood ratio test is used to produce a match score. If match score is larger than the decision threshold, the claimed speaker is accepted as a true user and otherwise. As PRP is inherited from the MRP formulation, we shall consider two performance scenarios that mentioned in section 2.

4 Experiments and Discussion

The experiments are conducted by using YOHO speech corpus [12] for text-independent speaker verification. The details of the database are shown in Table 1. The database consists of i=138 speakers and 13 samples each. In the experiment, 5 samples of each speaker are randomly selected for training while the others j=8 samples are used for testing. The speech signal is blocked into 240 speech samples per frame with 160 overlapped with adjacent frames. The Linear Predictive (LP) cepstrum with Hamming Window is used to extract the speech feature set. The dimension of speech feature set is qxr, where q is Linear Predictive Coefficients (LPC) order which is fixed to 30 and r refers to the number of frames. Gaussian mixture order is set to 20. Universal Background Model (UBM) approach is adopted to build the speaker background model. The resulting log scores are normalized so that they are within the range [0 1].

Table 1. Database used in the experiments

Number of speakers	138
Male : Female	106 : 32
Training speech sample per speaker	5
Test speech sample per speaker	8
Average duration of speech sample	4 sec per speech sample
Sampling frequency	8kHz

The system performance will be evaluated by using False Accept Rate (FAR), False Reject Rate (FRR) and Equal Error Rate (EER). The EER is obtained when the FAR and the FRR are equal. FRR is the rate in which the system incorrectly rejects a valid verification attempt and FAR is the rate in which the system wrongly accepts an invalid attempt.

For the imposter distribution in the FAR calculation, the first speech feature in the test set is fed into the speaker model, λ of all other speakers, and the same process is repeated for subsequent speech feature. Thus, it yields a total of 151248 (ix(i-1)xj) imposter probabilistic scores. For the genuine speaker distribution in the FRR test, each speech feature is fed into their corresponding speaker model, λ leading to 1104 (ixj) genuine probabilistic scores. We repeat the same process 20 times and the results are averaged to reduce the statistical frustration caused by the different random numbers.

We fix 2D PCA row dimension, p=30. In this paper, PRP-m and PRPs-m denotes PRP in genuine-token and stolen-token scenarios, respectively with various $m \leq p$ row dimensions (5, 10, 15, 20, 25, 27, 29 and 30). Table 2 shows the performance comparisons of 2DPCA-GMM, PRP-m, PRPs-m for various m. It is clearly shown that the PRP significantly outperforms the original method (2DPCA-GMM in this context). PRP attains zero EER when m increases beyond m=10, as what we expected to see in all user-specific token mixing algorithms [4][7].

For the stolen-token case, we take the worst scenario where the imposters always manage to steal the genuine token. In other words, only one set of PRN is applied to all speech samples and the feeding is done according to the imposter match described above. In Table 2, we observe the degraded performance of PRPs-*m* as compared to PRP-*m*. However, PRPs-*m* with EER=5.96% (*m*=29) and EER=6.43 % (*m*=27) are close to 2DPCA-GMM with EER=5.78% when *m* is slightly less than *p*. In other words, PRPs-*m* reverts the system to its original state when *m*≈*p*. From Table 3, it is showed that the imposter distribution of mean and variance (0.5640 and 0.0110) for PRPs-*m* are close to the 2DPCA-GMM with mean and variance (0.5628 and 0.0118), when *m*=29. Similar result has been seen in the genuine distribution where the mean and variance (0.7880 and 0.0055) of PRPs-*m* are close to the mean and variance of 2DPCA-GMM (0.7937 and 0.0057), when *m*=29. This indicates that the preservation of genuine-imposter distribution is valid when *m*≈*p*. By setting the *m* slightly less than *p*, the performance will be retained and this does not jeopardize the condition of non-invertibility. Similar result is also presented in MRP [7] which utilizes the normalized dot-product as the matching metric. In PRP, we employ the probabilistic score, which is derived from GMM. Although different matching techniques are employed, both methods lead to preservation of the intra-class variations (genuine distribution) as well as inter-class variations (imposter distribution). It is also depicted in Fig. 3 where the separation of the genuine-imposter class distribution for 2DPCA-GMM and PRPs-*m* are almost identical.

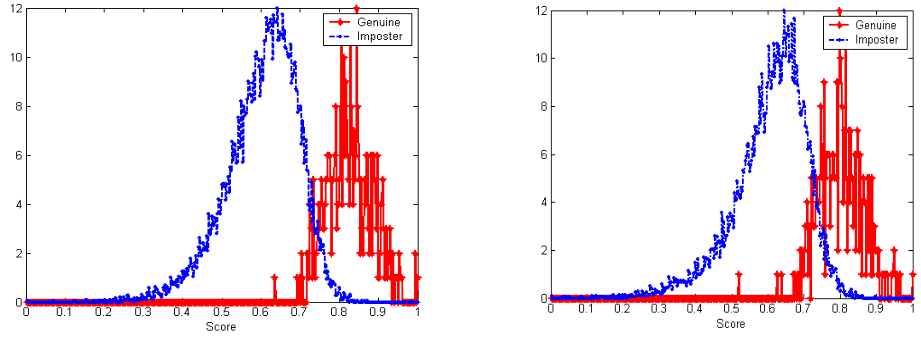
To fulfill the diversity requirement of cancellable biometrics, we examine whether the PRP with PRN A and PRP with PRN B (both with same speech feature) are associated. This can be done by using Pairwise Independent Test. We mix the same speech feature with different PRN and the scores generation procedure is followed

Table 2. Performance comparison for 2DPCA-GMM, PPR-*m*, PRPs-*m*

	m	FAR(%)	FRR(%)	EER(%)
2DPCA-GMM	-	5.76	5.79	5.78
PRP-*m*	5	1.65	1.63	1.64
	10	0.03	0	0.02
	15	0	0	0
	20	0	0	0
	25	0	0	0
	27	0	0	0
	29	0	0	0
	30	0	0	0
PRPs-*m*	5	15.89	15.94	15.92
	10	12.85	12.95	12.90
	15	10.33	10.41	10.37
	20	8.76	9.10	8.93
	25	7.80	8.11	7.96
	27	6.12	6.73	6.43
	29	5.81	6.11	5.96
	30	5.69	5.70	5.70

Table 3. Statistic measurement for 2DPCA-GMM and PRPs-*m* (stolen-token scenario)

	m	μ_g	μ_i	σ_g^2	σ_i^2
2DPCA-GMM	-	0.7937	0.5628	0.0057	0.0118
PRPs-*m*	5	0.8205	0.7054	0.0031	0.0066
	10	0.7700	0.5794	0.0054	0.0105
	15	0.7784	0.5674	0.0055	0.0110
	20	0.8002	0.5842	0.0051	0.0107
	25	0.7895	0.5680	0.0054	0.0109
	27	0.7787	0.5736	0.0053	0.0109
	29	0.7880	0.5640	0.0055	0.0110
	30	0.7964	0.5620	0.0058	0.0121



(a)

(b)

Fig. 3. Genuine and Imposter class distribution for (a) 2DPCA-GMM and (b) PRPs-27 (Probabilistic scores were normalized to [0-1] for better visualization purpose)

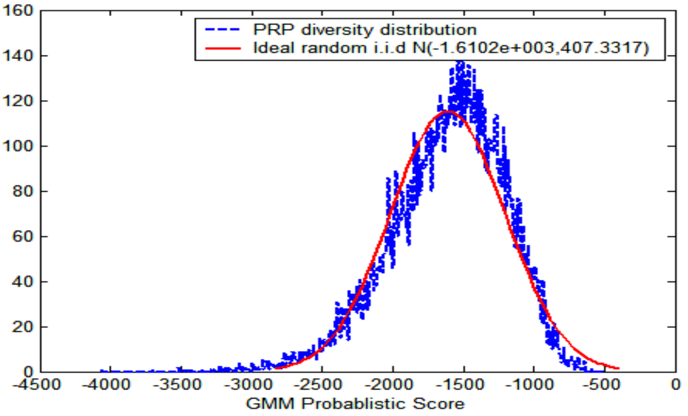


Fig. 4. Pairwise Independent Test of PRP

exactly by the imposter scores collection as described above. As shown in Fig. 4, the mean and standard deviation of collected scores are -1483.4 and 373.1, respectively. As the histogram closely approaches the independent and identically distributed (i.i.d) random variables drawn from Gaussian distribution, $\mathcal{N}$(-1610.2, 407.3), we can conclude that the PRP is pairwise independent. This implies that the refreshed PRP has almost no correlation with old PRP, and hence random number refreshment is equivalent to issue a new template for the user.

From the above findings, we observed that PRP fulfills the requirement of cancellable biometrics in term of performance, even in stolen-token scenario whereby the performance is retained as at feature vector level. In practical usage, we should set the system threshold t, which is used to decide the acceptance/rejection of the users according to the feature vector level performance (or stolen-token performance profile), instead of other scenarios. Nevertheless, recall that our results are important contribution to preserving the privacy of the speech feature and enable the enrolled template to be replaced in the event of template compromise.

5 Conclusions and Future Works

In this paper, we proposed a cancellable biometrics formulation, coined as Probabilistic Random Projections (PRP) which extends the MRP by employing the 2D Principal Component Analysis Gaussian Mixture Model. The PRP represents the speech feature in the 2D matrix format and hence, removes the limitation of MRP, which is only applicable to 1D fixed length feature vector. Besides that, the probabilistic modeling leads to better performance in speaker verification. The projection is non-invertible. The irrevocable and non-replacement issue of biometrics can be solved through the PRN replacement once the biometric template is compromised. Experiments showed that PRP survives in stolen-token attacks when the random subspace dimension is near to the original feature dimension. Thus, PRP functions well without compromising the verification performance in the event of compromised token. We also showed that PRP fulfilled another important property of cancellable biometrics, ie. diversity property. Our future work will be focusing on the theoretical justification on why the statistical preservation transformation can be achieved through the probabilistic scores that derived from GMM.

Acknowledgement

This work was supported by the Korea Science and Engineering Foundation (KOSEF) through the Biometrics Engineering Research Center (BERC) at Yonsei University.

References

1. Ratha, N., Connell, J., Bolle, R.M.: Enhancing Security and Privacy in Biometrics-based Authentication Systems. IBM Syst. J. 40, 614–634 (2001)
2. Uludag, U., Pankanti, S., Prabhakar, S., Jain, A.K.: Biometric Cryptosystems: Issues and Challenges. Proc. IEEE 92, 948–960 (2004)

3. Soutar, C., Roberge, D., Stoianov, A.R., Gilroy, Vijaya Kumar, B.V.K.: Biometrics Encryption. In: Nichols, R.K (ed.) ICSA Guide to Cryptography, pp. 649–675. McGraw-Hill, New York (1999)
4. Teoh, B.J.A., Ngo, C.L.D., Goh, A.: Personalised Cryptographic Key Generation Based on FaceHashing. Computers and Security J. 23, 606–614 (2004)
5. Savvides, M., Vijava Kumar, B.V.K., Khosla, P.K.: Cancelable Biometrics Filters for Face Recognition. Int. Conf. of Pattern Recognition 3, 922–925 (2004)
6. Chong, T.Y., Teoh, B.J.A., Ngo, C.L.D., Goh, M.: Multi-Space Random Mapping for Speaker Identification. IEICE Electron. Express 2, 226–231 (2005)
7. Teoh, B.J.A., Ngo, C.L.D., Goh, A.: Random Multispace Quantisation as an Analytic Mechanism for BioHashing of Biometric and Random Identity Inputs. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(12), 1892–1901 (2006)
8. Chong, L.Y., Teoh, B.J.A., Khor, S.E.: 2D CLAFIC Subspace Technique in Probabilistic Speaker Verification. In: Fifth IEEE Workshop on Automatic Identification Advanced Technologies. IEEE Computer Society Press, Los Alamitos (2007)
9. Teoh, B.J.A., Chong, T.Y.: Cancellable Biometrics Realization with Multispace Random Projections. Accepted to publish in IEEE Transactions SMC Part B (2007)
10. Makhoul, J.: Linear Prediction: A tutorial review. Proc. IEEE 63, 561–580 (1975)
11. Reynolds, D.A.: Speaker Verification using Adapted Gaussian Mixture Models. Digital Signal Processing 10, 19–41 (2000)
12. Higgins, A.: YOHO Speaker Verification. In: Speech Research Symposium (1990)