# A GMM SUPERVECTOR KERNEL WITH THE BHATTACHARYYA DISTANCE FOR SVM BASED SPEAKER RECOGNITION

Chang Huai You, Kong Aik Lee, Haizhou Li

Institute for Infocomm Research (I<sup>2</sup>R), A\*STAR, 1 Fusionopolis Way, Singapore Emails: {echyou, kalee, hli}@i2r.a-star.edu.sg

### ABSTRACT

Gaussian mixture model (GMM) supervector is one of the effective techniques in text independent speaker recognition. In our previous work, we introduce the GMM-UBM mean interval (GUMI) concept based on the Bhattacharyya distance. Subsequently GUMI kernel was successfully used in conjunction with support vector machine (SVM) for speaker recognition. Besides the first order statistics, it is generally believed that speaker cues are also partly conveyed by second order statistics. In this paper, we extend the Bhattacharyya-based SVM kernel by constructing the supervector with the mean statistical vector and the covariance statistical vector. Comparing with the Kullback-Leibler (KL) kernel, we demonstrate the effectiveness of the new kernel on the 2006 National Institute of Standards and Technology (NIST) speaker recognition evaluation (SRE) dataset.

*Index Terms*— Gaussian Mixture Model, Support Vector Machine, Supervector, Speaker Verification, NIST Evaluation

## 1. INTRODUCTION

Speaker recognition is usually formulated as a hypothesis test that verifies an identity claim by estimating the similarity of the claimant's speech and the enrolled utterance(s). In text-independent speaker verification, both Gaussian mixture model (GMM) and support vector machine have been proven to be effective and most popularly used for many years.

In [1], Campbell *et al.* constructed an SVM kernel using the GMM supervector that is formed by using the parameters of GMM. They derive an SVM kernel based on KL divergence between two GMMs. However, only the adapted means are considered while some information carried by covariance is ignored. This kernel is also suitable for language recognition. In [2], the kernel is extended by using the symmetrized version of the KL divergence for language recognition. As a result, the covariance term can be introduced into the kernel. In other words, the extended KL-based kernel measures the similarity in terms of mean and covariance.

The Bhattacharyya distance [3] has been noticed to have several applications in classical statistics, it turns out to give better results than the divergence. In our recent study [4], we introduced a mean interval concept based on the Bhattacharyya distance. Consequently, we proposed a core distance measurement for the similarity and derive a kernel with GMM-UBM mean interval (GUMI) supervector. In GMM, covariance is estimated for the purpose of carrying some useful information about speaker. It is reasonable to infer that the covariance term in the derived Bhattacharyya distance of two GMMs may reveal some underlaying speaker information. Based on the above motivation, in this paper, we extend the GUMI concept to include not only the Bhattacharyya mean statistical term but also the covariance term. We give the derivation of the new kernel based on the Bhattacharyya distance.

We compare the proposed kernel with the conventional linear KL kernel under exactly same experimental conditions and follows the testing task conducted by NIST SRE 2006 evaluation. In the performance evaluation, we also apply the strictly same implementation procedure of nuisance attribute projection (NAP) for both conventional and proposed kernels. The NAP is done by removing the nuisance subspace from the GMM supervector by projection. Through the experiments, it is observed that the proposed kernel has a consistently superior performance over the conventional KL kernel in terms of equal error rate (EER), minimum detection cost function (minDCF) and detection error trade-off (DET). In the remainder of the paper, we introduce the conventional KL kernel in section 2. We propose the Bhattacharyya based kernel for GMM-supervector SVM in section 3. The performance evaluation is shown in section 4. Finally we summarize the paper and give a discussion in section 5.

## 2. CONVENTIONAL GMM-SVM SYSTEM

#### 2.1. KL Divergence Kernel

GMM-supervector SVM combines both generative and discriminative methods and leads to the generative SVM kernels based on the probability distribution estimation. Recently, GMM-supervector kernel [1] becomes state-of-the-art approach in speaker recognition. In conventional GMMsupervector SVM system, KL divergence is used to measure the distance of the two GMMs. The KL divergence between the two distributions with probability density  $p_a$  and  $p_b$  is given by

$$\Psi_{\rm KL}(p_a||p_b) = \int_{R^n} p_a(\mathbf{x}) \log\left(\frac{p_a(\mathbf{x})}{p_b(\mathbf{x})}\right) d\mathbf{x}.$$
 (1)

With the definition in (1), the KL divergence for two Gaussian probability distributions is obtained by

$$\Psi_{\rm KL}(f(\mathbf{m}_i^{(a)}, \Sigma_i^{(a)}) || f(\mathbf{m}_i^{(b)}, \Sigma_i^{(b)})) = \frac{1}{2} \Big\{ \ln(\frac{|\Sigma_i^{(b)}|}{|\Sigma_i^{(a)}|}) + tr((\Sigma_i^{(b)})^{-1} \Sigma_i^{(a)}) + (\mathbf{m}_i^{(b)} - \mathbf{m}_i^{(a)})^T (\Sigma_i^{(b)})^{-1} (\mathbf{m}_i^{(b)} - \mathbf{m}_i^{(a)}) - D \Big\}$$
(2)

where  $\mathbf{m}_i$  and  $\Sigma_i$  denote the respective mean and covariance of the *i*-th Gaussian component of the GMM a or b; and *f* represents the Gaussian distribution function. Since the KL divergence does not satisfy the Mercer condition, for GMM distribution, unchange of the weight should be assumed, i.e.  $\omega_i^{(u)} = \omega_i^{(a)} = \omega_i^{(b)}$ , where u denotes UBM. On the other side, an approximation for two GMM distance is considered by bounding the divergence with the log-sum inequality

$$\Psi_{\rm KL}(p_a||p_b) \leq \sum_{i=1}^{M} \Psi_{\rm KL}(p_{a_i}||p_{b_i})$$
  
=  $\sum_{i=1}^{M} \Psi_{\rm KL}(\omega_i^{(u)} f(\mathbf{m}_i^{(a)}, \Sigma_i^{(a)})||\omega_i^{(u)} f(\mathbf{m}_i^{(b)}, \Sigma_i^{(b)}))$  (3)  
=  $\sum_{i=1}^{M} \omega_i^{(u)} \Psi_{\rm KL}(f(\mathbf{m}_i^{(a)}, \Sigma_i^{(a)})||f(\mathbf{m}_i^{(b)}, \Sigma_i^{(b)}))$ 

where M is the number of Gaussian components in the GMM. With the assumption that the covariance adaptation can be ignored, i.e.  $\Sigma_i^{(u)} = \Sigma_i^{(a)} = \Sigma_i^{(b)}$ , the linear kernel function can be given as follows [5]

$$K_{\rm KL}(\mathbf{X}_a, \mathbf{X}_b) = \sum_{i=1}^{M} (\sqrt{\omega_i^{(u)}} (\Sigma_i^{(u)})^{-\frac{1}{2}} \mathbf{m}_i^{(a)})^T (\sqrt{\omega_i^{(u)}} (\Sigma_i^{(u)})^{-\frac{1}{2}} \mathbf{m}_i^{(b)})$$
(4)

where  $\mathbf{X}_a$  and  $\mathbf{X}_b$  denote the respective feature vector sequences of utterances a and b.

## 2.2. A Covariance KL Kernel

Although the KL divergence can be used to represent the difference between two probability distributions, it is neither positive definite nor symmetric, so that it cannot be straightforwardly used for a kernel. In [2], the KL kernel is extended to introduce the covariance term by using the symmetrized version of the KL divergence. In other words, the extended KL kernel contains two terms, i.e., mean vector term and covariance term

$$K_{\text{extKL}}(\mathbf{X}_{a}, \mathbf{X}_{b}) = \sum_{i=1}^{M} (\sqrt{\omega_{i}^{(\mathbf{u})}} (\Sigma_{i}^{(\mathbf{u})})^{-\frac{1}{2}} \mathbf{m}_{i}^{(\mathbf{a})})^{T} (\sqrt{\omega_{i}^{(\mathbf{u})}} (\Sigma_{i}^{(\mathbf{u})})^{-\frac{1}{2}} \mathbf{m}_{i}^{(\mathbf{b})}) + \sum_{i=1}^{M} \frac{\omega_{i}^{(\mathbf{u})}}{2} tr(\Sigma_{i}^{(\mathbf{a})} (\Sigma_{i}^{(\mathbf{u})})^{-2} \Sigma_{i}^{(\mathbf{b})})$$
(5)

where  $tr(\cdot)$  denotes the trace of matrix.

## 3. PROPOSED BHATTACHARYYA-BASED GMM-SVM KERNEL

#### 3.1. The Bhattacharyya Distance of Two GMMs

The Bhattacharyya kernel (or Bhattacharyya coefficient) [3] [6] for two probability distributions is defined by

$$\Lambda_{\text{Bhatt}}(p_a||p_b) = \int_{R^n} \sqrt{p_a(\mathbf{x})} \sqrt{p_b(\mathbf{x})} \, d\mathbf{x} \tag{6}$$

and the Bhattacharyya distance of the two probability distributions is defined by

$$\Psi_{\text{Bhatt}}(p_a||p_b) = -\ln(\Lambda_{\text{Bhatt}}(p_a||p_b)).$$
(7)

Let  $p_a$  and  $p_b$  denote the two probability distributions of GMM<sub>a</sub> and GMM<sub>b</sub>. The Bhattacharyya distance for the two GMMs are given by

$$\Psi_{\text{Bhatt}}(p_a||p_b) = -\ln\left(\int_{R^n} \sqrt{\sum_{i=1}^M p_{a_i}(\mathbf{x})} \sqrt{\sum_{j=1}^M p_{b_j}(\mathbf{x})} \, d\mathbf{x}\right)$$
(8)

where  $p_{a_i} = \omega_i^{(a)} f(\mathbf{x} | \mathbf{m}_i^{(a)}, \Sigma_i^{(a)})$  and  $p_{b_j} = \omega_j^{(b)} f(\mathbf{x} | \mathbf{m}_j^{(b)}, \Sigma_j^{(b)})$ . The Bhattacharyya distance between the *i*-th Gauss of GMM<sub>a</sub> and the corresponding one of GMM<sub>b</sub> is obtained

$$\begin{split} \Psi_{\text{Bhatt}}(p_{a_{i}}||p_{b_{i}}) \\ &= -\ln\left[\int_{\mathbb{R}^{n}}\sqrt{\omega_{i}^{(a)}f(\mathbf{x}|\mathbf{m}_{i}^{(a)},\Sigma_{i}^{(a)})}\sqrt{\omega_{i}^{(b)}f(\mathbf{x}|\mathbf{m}_{i}^{(b)},\Sigma_{i}^{(b)})} \, d\mathbf{x}\right] \\ &= \frac{1}{8}(\mathbf{m}_{i}^{(b)} - \mathbf{m}_{i}^{(a)})^{T}\left[\frac{\Sigma_{i}^{(a)} + \Sigma_{i}^{(b)}}{2}\right]^{-1}(\mathbf{m}_{i}^{(b)} - \mathbf{m}_{i}^{(a)}) \\ &\quad + \frac{1}{2}\ln\frac{|\frac{\Sigma_{i}^{(a)} + \Sigma_{i}^{(b)}|}{2}|}{\sqrt{|\Sigma_{i}^{(a)}||\Sigma_{i}^{(b)}|}} - \frac{1}{2}\ln(\omega_{i}^{(a)}\omega_{i}^{(b)}). \end{split}$$
(9)

Bounding (8) with the log-sum inequality, we have

$$\begin{split} \Psi_{\text{Bhatt}}(p_{a}||p_{b}) &= \Psi_{\text{Bhatt}}(\left[\sum_{i=1}^{M} p_{a_{i}}\right]||\left[\sum_{i=1}^{M} p_{b_{i}}\right]) \\ &\leq \sum_{i=1}^{M} \Psi_{\text{Bhatt}}(p_{a_{i}}||p_{b_{i}}) \\ &= \frac{1}{8}\sum_{i=1}^{M} \left\{ (\mathbf{m}_{i}^{(b)} - \mathbf{m}_{i}^{(a)})^{T} \left[\frac{\Sigma_{i}^{(a)} + \Sigma_{i}^{(b)}}{2}\right]^{-1} (\mathbf{m}_{i}^{(b)} - \mathbf{m}_{i}^{(a)}) \right\} \\ &+ \frac{1}{2}\sum_{i=1}^{M} \left[ \ln \frac{|\frac{\Sigma_{i}^{(a)} + \Sigma_{i}^{(b)}}{2}|}{\sqrt{|\Sigma_{i}^{(a)}||\Sigma_{i}^{(b)}|}} \right] + \frac{1}{2}\sum_{i=1}^{M} \ln \left\{ \left( \omega_{i}^{(a)} \omega_{i}^{(b)} \right)^{-1} \right\}. \end{split}$$
(10)

## 3.2. GUMI Kernel

On the right hand side of (10), there are three terms. The first term reflects the degree of mean statistical similarity; and the second represents the degree of consistency of the covariance matrices; the third term is the weighting factor. The first term contains the most informative quantity to measure the similarity of the two GMMs. In our previous work [4], we have introduced the GUMI kernel. According to the first term, i.e.  $\sum_{i=1}^{M} \left\{ (\mathbf{m}_{i}^{(b)} - \mathbf{m}_{i}^{(a)})^{T} \left[ \frac{\sum_{i=1}^{(a)} + \sum_{i=1}^{(b)}}{2} \right]^{-1} (\mathbf{m}_{i}^{(b)} - \mathbf{m}_{i}^{(a)}) \right\}, \text{ we derived the GUMI kernel [4]}$ 

$$K_{\text{GUMI}}(\mathbf{X}_{a}, \mathbf{X}_{b}) = \sum_{i=1}^{M} \left\{ \left[ \left( \frac{\Sigma_{i}^{(a)} + \Sigma_{i}^{(u)}}{2} \right)^{-\frac{1}{2}} (\mathbf{m}_{i}^{(a)} - \mathbf{m}_{i}^{(u)}) \right]^{T} \\ \left[ \left( \frac{\Sigma_{i}^{(b)} + \Sigma_{i}^{(u)}}{2} \right)^{-\frac{1}{2}} (\mathbf{m}_{i}^{(b)} - \mathbf{m}_{i}^{(u)}) \right] \right\}$$
(11)

#### 3.3. The New Kernel with the Bhattacharyya Distance

From (10), the Bhattacharyya distance between the two GMMs can be constrained as follows

$$\begin{split} \breve{\Psi}_{\text{Bhatt}}(p_{a}||p_{b}) \\ &= \frac{1}{8} \sum_{i=1}^{M} \left\{ (\mathbf{m}_{i}^{(b)} - \mathbf{m}_{i}^{(a)})^{T} \left[ \frac{\Sigma_{i}^{(a)} + \Sigma_{i}^{(b)}}{2} \right]^{-1} (\mathbf{m}_{i}^{(b)} - \mathbf{m}_{i}^{(a)}) \right\} \\ &+ \frac{1}{2} \sum_{i=1}^{M} \left[ \ln \frac{|\frac{\Sigma_{i}^{(a)} + \Sigma_{i}^{(b)}|}{\sqrt{|\Sigma_{i}^{(a)}||\Sigma_{i}^{(b)}|}} \right] - \frac{1}{2} \sum_{i=1}^{M} \ln(\omega_{i}^{(a)} \omega_{i}^{(b)}) \\ &\approx \frac{1}{8} \sum_{i=1}^{M} \left\{ (\mathbf{m}_{i}^{(b)} - \mathbf{m}_{i}^{(a)})^{T} \left[ \frac{\Sigma_{i}^{(a)} + \Sigma_{i}^{(b)}}{2} \right]^{-1} (\mathbf{m}_{i}^{(b)} - \mathbf{m}_{i}^{(a)}) \right\} \\ &+ \frac{1}{2} \sum_{i=1}^{M} \frac{|\frac{\Sigma_{i}^{(a)} + \Sigma_{i}^{(b)}|}{\sqrt{|\Sigma_{i}^{(a)}||\Sigma_{i}^{(b)}|}} - \frac{M}{2} + \sum_{i=1}^{M} \ln \left\{ \frac{1}{\sqrt{\omega_{i}^{(a)} \omega_{i}^{(b)}}} \right\} \end{split}$$
(12)

In (12), we use  $\ln(x) \approx x - 1$ , while  $x \to 1$ . Since  $\sum_{i}^{(\lambda)}$ ( $\lambda$  represents a or b) is adapted from  $\sum_{i}^{(u)}$ , thus  $\frac{|\frac{\sum_{i=2}^{(a)} + \sum_{i=1}^{(b)}}{\sqrt{|\sum_{i}^{(a)}||\sum_{i}^{(b)}|}}$ approaches 1. As a result,  $\ln \frac{|\frac{\sum_{i=2}^{(a)} + \sum_{i=1}^{(b)}}{\sqrt{|\sum_{i}^{(a)}||\sum_{i}^{(b)}|}} \approx \frac{|\frac{\sum_{i=2}^{(a)} + \sum_{i=1}^{(b)}}{\sqrt{|\sum_{i}^{(a)}||\sum_{i}^{(b)}|}} - 1$ . According to the GUMI analysis [4], we may reach an approximation of the equation (12)

$$\begin{split} \breve{\Psi}_{\text{Bhatt}}(p_{a}||p_{b}) \\ \approx \frac{1}{8} \sum_{i=1}^{M} \left\{ \left[ \left( \frac{\Sigma_{i}^{(a)} + \Sigma_{i}^{(u)}}{2} \right)^{-\frac{1}{2}} (\mathbf{m}_{i}^{(a)} - \mathbf{m}_{i}^{(u)}) \right]^{T} \\ \left[ \left( \frac{\Sigma_{i}^{(b)} + \Sigma_{i}^{(u)}}{2} \right)^{-\frac{1}{2}} (\mathbf{m}_{i}^{(b)} - \mathbf{m}_{i}^{(u)}) \right] \right\} \\ + \frac{1}{2} \sum_{i=1}^{M} tr \left[ \left( \frac{\Sigma_{i}^{(a)} + \Sigma_{i}^{(u)}}{2} \right)^{\frac{1}{2}} (\Sigma_{i}^{(a)})^{-\frac{1}{2}} \left( \frac{\Sigma_{i}^{(b)} + \Sigma_{i}^{(u)}}{2} \right)^{\frac{1}{2}} (\Sigma_{i}^{(b)})^{-\frac{1}{2}} \right. \\ + \sum_{i=1}^{M} \ln \left\{ \frac{1}{\sqrt{\omega_{i}^{(a)} \omega_{i}^{(b)}}} \right\} - \frac{M}{2} \end{split}$$
(13)

Considering the first two terms of (13), we propose the kernel as follows

$$\begin{split} K_{\text{Bhatt2}}(\mathbf{X}_{a}, \mathbf{X}_{b}) \\ &= \sum_{i=1}^{M} \left\{ \left[ \frac{1}{2} \left( \frac{\Sigma_{i}^{(a)} + \Sigma_{i}^{(u)}}{2} \right)^{-\frac{1}{2}} (\mathbf{m}_{i}^{(a)} - \mathbf{m}_{i}^{(u)}) \right]^{T} \\ & \left[ \frac{1}{2} \left( \frac{\Sigma_{i}^{(b)} + \Sigma_{i}^{(u)}}{2} \right)^{-\frac{1}{2}} (\mathbf{m}_{i}^{(b)} - \mathbf{m}_{i}^{(u)}) \right] \right\} \\ &+ \sum_{i=1}^{M} tr \left[ \left( \frac{\Sigma_{i}^{(a)} + \Sigma_{i}^{(u)}}{2} \right)^{\frac{1}{2}} (\Sigma_{i}^{(a)})^{-\frac{1}{2}} \left( \frac{\Sigma_{i}^{(b)} + \Sigma_{i}^{(u)}}{2} \right)^{\frac{1}{2}} (\Sigma_{i}^{(b)})^{-\frac{1}{2}} \right] \end{split}$$
(14)

In this case, the GMM supervector is the concatenation of the mean statistical vector and the covariance vector. Apparently, the kernel is an inner product of the two supervectors marked by a and b. It satisfies the Mercer condition [7] and suitable for SVM. The SVM is a two-class classifier making use of hyperplane separator, which is estimated by maximizing the distance between the hyperplane and the closest training vectors, e.g. the GMM supervectors, which are called support vectors.

## 4. PERFORMANCE EVALUATION

#### 4.1. Evaluation Configuration

The performance evaluation is conducted on the NIST SRE 2006 core test, where 51448 trials are tested, which includes 3612 true trials and 47836 false trials. We carried out gender-dependent speaker verification experimentation. 36-dimension linear predictive cepstral coefficient (LPCC) feature and 512 Gaussian mixtures of GMM are selected.



Fig. 1. Comparison of the GMM-Supervector SVM (GS-SVM) Systems with different kernels by using gender-dependent testing for the 51448 trials (including male and female) of NIST SRE-2006 1conv4w-1conv4w task. 'extKL' denotes the extended KL kernel by (5), and 'Bhatt2' denotes our proposed Bhattacharyya-based kernel by (14).

Thus, the length of supervector for the conventional KL kernel (refer to equation (4)) and the GUMI kernel (refer to equation (11)) is  $512 \times 36 = 18432$ ; while the length of supervector for the extended KL kernel (refer to equation (5)) and the proposed Bhattacharyya-based kernel (14) is  $512 \times 36 \times 2 = 36864$ . The UBM is trained through EM algorithm by using NIST 2004 1-side training database. The GMM-supervector is obtained using MAP adaptation for each utterance.

During the entire GMM-supervector performance evaluation, the conventional KL kernel ((4), [1]), extended KL kernel ((5), [2]), the GUMI kernel ((11), [4]) and the new proposed Bhattacharyya-based kernel (14) are implemented in the way to have exactly same training and testing conditions such as feature extraction processing, feature enhancement, background databases for both GMM and SVM, training and testing databases, test-normalization (T-Norm) models, NAP training database and its parameter configuration. The SRE 2005 are used as cohort models for T-Norm. For the case of gender-dependent T-Norm, 271 NIST SRE 2005 models for male and 362 models for female are used for the respective male and female trials.

Fig. 1 plots the DET curves with different GMM-SVM kernels where NIST 2004 training database used for the NAP [8] and SVM background. Table 1 shows EERs and minimum DCFs corresponding to the DET curves in Fig. 1. From the experimental results, it is observed that the performance by using the proposed kernel is better than those by using GUMI kernel as well as the two conventional KL kernels.

In this paper, we only show the linear kernel comparison. The linear kernel is the scalar product of the supervectors depending on the approximation of the KL divergence or Bhattacharyya distance, whereas the nonlinear kernel could be connected to an exponential function of the supervectors according to the distance approximation. Actually, the GMM supervector kernels reflected in (4), (5), (11) and (14) can be extended to the nonlinear kernels. It would be an interesting work to do further comparison for the nonlinear kernels.

**Table 1**. The comparison of the Equal Error Rates and minimum DCFs of the various GMM-SVM kernels by using the 512-mixtures of GMM on the SRE-2006 1conv4w-1conv4w evaluation, 51448 trials.

Gender-depend.	EER & minDCF(x100)	
SRE06 Core Test	Raw-Score	Score with S05
		T-Norm
NAP + KL	6.45% & 3.31	5.51% & 2.69
NAP + extKL	6.31% & 3.35	5.57% & 2.82
NAP + GUMI	6.20% & 3.04	5.36% & 2.62
NAP + Bhatt2	5.95% & 2.95	5.12% & 2.48

#### 5. SUMMARY AND DISCUSSION

In this paper, we apply the Bhattacharyya distance to measure the similarity between two GMM distributions. We extend the GUMI kernel to involve the covariance term and propose a new Bhattacharyya-based kernel, which satisfies the Mercer kernel condition. Therefore the SVM can be used in the new kernel. The extended Bhattacharyya-based kernel gives improvement as compared to the GUMI and the two conventional KL kernels through experiment on the NIST SRE 2006 core testing task.

#### 6. REFERENCES

- W. M. Campbell, D. E. Sturim, and D. A. Reynolds, "Support vector machines using GMM supervectors for speaker verification," *IEEE Signal Processing Letters*, Vol. 13, pp. 308-311, 2006.
- [2] W. M. Campbell, "A Covariance kernel for SVM language recognition," *Int. Conf. Acoust. Speech and Signal Process.*, pp. 4141-4144, 2008.
- [3] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans. Commun. Technol.*, Vol. COM-15. pp. 52-60, Feb. 1967.
- [4] C. H. You, K. A. Lee and H. Li, "An SVM kernel with GMMsupervector based on the Bhattacharyya distance for speaker recognition," *IEEE Signal Processing Letters*, vol. 16, no. 1, pp. 49-52, Jan. 2009.
- [5] W. M. Campbell, D. E. Sturim, D. A. Reynolds, and A. Solomonoff, "SVM based speaker verification using a GMM supervector kernel and NAP variability compensation," *Int. Conf. Acoust. Speech and Signal Process.*, pp. 97-100, France, 2006.
- [6] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bull. Calcutta Math. Soc.*, Vol. 35, pp. 99-109, 1943.
- [7] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge, 2004.
- [8] A. Solomonoff, W. M. Campbell, and C. Quillen, "Channel compensation for SVM speaker recognition," *Proc. Odyssey04*, pp. 57-62, 2004.