Cosine Similarity Scoring without Score Normalization Techniques

Najim Dehak[1] , Reda Dehak[2], James Glass[1], Douglas Reynolds[3], Patrick Kenny[4]

[1] MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA USA
[2] Laboratoire de Recherche et de Dveloppement de l'EPITA (LRDE), Paris, France
[3] MIT Lincoln Laboratory, Lexington, MA USA
[4] Centre de Recherche d'Informatique de Montral (CRIM), Montral CANADA

najim@csail.mit.edu, reda@lrde.epita.fr, glass@mit.edu, dar@ll.mit.edu
patrick.kenny@crim.ca

Abstract

In recent work [1], a simplified and highly effective approach to speaker recognition based on the cosine similarity between low-dimensional vectors, termed *ivectors*, defined in a total variability space was introduced. The total variability space representation is motivated by the popular Joint Factor Analysis (JFA) approach, but does not require the complication of estimating separate speaker and channel spaces and has been shown to be less dependent on score normalization procedures, such as z-norm and t-norm. In this paper, we introduce a modification to the cosine similarity that does not require explicit score normalization, relying instead on simple mean and covariance statistics from a collection of impostor speaker ivectors. By avoiding the complication of z- and t-norm, the new approach further allows for application of a new unsupervised speaker adaptation technique to models defined in the ivector space. Experiments are conducted on the core condition of the NIST 2008 corpora, where, with adaptation, the new approach produces an equal error rate (EER) of 4.8% and min decision cost function (MinDCF) of 2.3% on all female speaker trials.

1. Introduction

Over recent years, Joint Factor Analysis (JFA) has demonstrated state of the art performance for text-independent speaker detection tasks in the NIST speaker recognition evaluations (SREs). JFA proposes powerful tools to enhance classic Gaussian Mixture Model (GMM) speaker models to represent speaker variability and to compensate for channel/session variability. However, JFA produces highly variable scores that require application of score normalization techniques, such as zt-norm, to show performance gains [2]. Recently, motivated by JFA techniques, a simplified and highly effective approach to speaker recognition based on the cosine similarity between low-dimensional vectors, termed *ivectors*, defined in a total variability space was introduced [1]. This total variability approach, avoids joint estimation of separate speaker and session spaces and factors, and is less reliant on application of score normalizations [3, 1].

In this paper we introduce a modification to the cosine similarity that does not require explicit score normalization, relying instead on simple mean and covariance statistics from a collection of impostor speaker ivectors. We derive a new scoring function that approximates the application of zt-norm. We further apply the new scoring to an unsupervised speaker adaptation algorithm proposed in [4], where the new scoring greatly simplifies the adaptation since there is no need to update score normalization parameters.

This paper is organized as follows. In Section 1.1, we describe the total variability space and the original cosine similarity scoring. Intersession compensation techniques used in the total variability space are described in Section 2. In Section 3 we analyze z- and t-norm score normalization in the context of the cosine similarity and, in Section 4, propose a modification to the cosine similarity scoring that approximates these score normalizations. Results using the new scoring on the 2008 NIST SRE corpus for both un-adapted and adapted approaches are presented in Section 5. Discussion and conclusions are given in Section 6.

1.1. Total variability

In JFA [2], a speaker utterance is represented by a supervector[2] (M) that consists of additive components from a speaker and a channel/session subspace. Specifically, the speaker-dependent supervector is defined as

$$M = m + Vy + Ux + Dz \tag{1}$$

where m is a speaker- and session-independent supervector (generally from a Universal Background Model (UBM)), V and D define a speaker subspace (eigenvoice matrix and diagonal residual, respectively), and U defines a session subspace (eigenchannel matrix). The vectors y, z and x are the speaker and session dependent factors in the respective subspaces and each is assumed to be a random variable with a Normal distribution $N(0, I)$. To apply JFA to speaker recognition consists of first estimating the subspaces (i.e., V, D, U) from appropriately labelled development corpora and then estimating the speaker-dependent factors (i.e., x, y, z) for each speaker and session. Scoring is done by computing the likelihood of the test utterance feature vectors against a session compensated speaker model ($M - Ux$). A comparison between several JFA scoring is given in [5].

In a more recent approach motivated by JFA [1], a speaker supervector is represented by factors in a single total variability space with no distinction made between speaker and session

[1]This work was sponsored by the Department of Defense under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

[2]A supervecvtor is composed by stacking the mean vectors from a GMM.

subspaces. The new speaker- and session- dependent supervector M is defined as

$$M = m + Tw \tag{2}$$

where m is the speaker- and session-independent supervector, the matrix T defines the total variability space, and the vector w is the speaker- and session-dependent factors in the total variability space (a random vector having a Normal distribution $N(0, I)$). Thus M is assumed to be normally distributed with mean vector m and covariance matrix TT^*. The process of training the total variability matrix T is equivalent to learning the eigenvoice V matrix [2], except for one important difference: in the eigenvoices training, all the recordings of a given speaker are considered to belong to the same person; in the case of the total variability matrix however, a given speaker's entire set of utterances are regarded as having been produced by different speakers. This new model can be viewed like a principal component analysis of the larger supervector space that allows us to project the speech utterances onto the total variability space. In this new speaker modeling, factor analysis plays the role of features extraction where we now operate on the total factor vectors,$\{w\}$, also called *ivectors*. Since the total variability space is significantly smaller then the supervector space (e.g., 400 vs. 122,880 dimensions), manipulations, such as compensations, modeling and scoring, become considerably more tractable.

1.2. Cosine similarity scoring

In the ivector space, a simple cosine similarity has been applied successfully to compare two utterances for making a speaker detection decision [6, 3]. Given two ivectors generated via projection of two supervectors into the total variability space, a target, w_{target}, from a known speaker and a test, w_{test}, from an unknown speaker, the cosine similarity score is given as

$$\text{score}\left(w_{\text{target}}, w_{\text{test}}\right) = \frac{\left(w_{\text{target}}\right)^t w_{\text{test}}}{\left\|w_{\text{target}}\right\| \cdot \left\|w_{\text{test}}\right\|} \gtrless \theta \tag{3}$$

where θ is the decision threshold. This scoring function is considerably less complex than scoring operations used in JFA [5]. Note that the cosine similarity only considers the angle between the two ivectors and not their magnitudes. It is believed that non-speaker information (such as session and channel) affects the ivector magnitudes so removing magnitude in scoring greatly improves the robustness of the ivector system.

2. Intersession compensation

In the total variability representation, there is no explicit compensation for inter-session variability as there is in JFA. However, once data has been projected into the low-dimensional total variability space it is rather straight forward and computationally efficient to apply standard compensation techniques. In [3], we tested three channel/session compensation techniques: Linear Discriminant Analysis (LDA), Nuisance Attribute Projection (NAP) and Within Class Covariance Normalization (WCCN). The best performance was obtained with the combination of LDA followed by WCCN. We next describe these compensations as applied to ivectors.

2.1. Linear Discriminant Analysis

LDA seeks to find a new orthogonal basis (rotation) of the feature space to better discriminate between different classes. Here, our classes are speakers. The new basis is sought to simultaneously maximize the between class variance (inter speaker discrimination) and minimize within class variance (intra speaker variability). These axes can be defined using a projection matrix A composed of the best eigenvectors (those with highest eigenvalues) of the general eigenvalues equation

$$\Sigma_b v = \lambda \Sigma_w v \tag{4}$$

where λ is the diagonal matrix of eigenvalues. The matrices Σ_b and Σ_w correspond to the between classes and within class covariance matrices, respectively. These are calculated as follows:

$$\Sigma_b = \sum_{i=1}^{S} \left(w_i - \overline{w}\right)\left(w_i - \overline{w}\right)^t \tag{5}$$

$$\Sigma_w = \sum_{s=1}^{S} \frac{1}{n_s} \sum_{i=1}^{n_s} \left(w_i^s - \overline{w_s}\right)\left(w_i^s - \overline{w_s}\right)^t \tag{6}$$

where $\overline{w_s} = \frac{1}{n_s}\sum_{i=1}^{n_s} w_i^s$ is the mean of the ivectors for each speaker, S is the total number of speakers and n_s is the number of utterances for each speaker s. In our previous work [6, 3], we assumed the mean vector of the entire speaker population $\overline{w}$ is equal to the null vector since the factors have a standard Normal distribution, $w \sim N(0, I)$, with zero mean. However, in more recent experiments, we used the actual computed global mean, rather than assuming it was zero, and found a slight performance improvement.

2.2. Within class covariance normalization

In [1], we successfully applied WCCN [7] as compensation to the ivector system, with the best performance obtained when it was preceded by LDA. The idea behind WCCN is to scale the ivector space inversely proportional to an estimate of the in-class covariance matrix, so that directions of high intra-speaker variability are deemphasized in ivector comparisons. The within class covariance is estimated using ivectors from a set of development speakers as

$$W = \frac{1}{S}\sum_{s=1}^{S} \frac{1}{n_s} \sum_{i=1}^{n_s} \left(A^t w_i^s - \overline{w_s}\right)\left(A^t w_i^s - \overline{w_s}\right)^t \tag{7}$$

where A is the LDA projection matrix, $\overline{w_s} = \frac{1}{n_s}\sum_{i=1}^{n_s} A^t w_i^s$ is the mean of the LDA projected ivectors for each speaker s , S is the total number of speakers, and n_s is the number of utterances of each speaker s. We use the inverse of this matrix in order to normalize the direction of the projected ivector components, which is equivalent to scaling the space by the matrix B, where $BB^t = W^{-1}$. With LDA and WCCN applied, the cosine similarity becomes

$$\text{score}\left(w_{\text{target}}, w_{\text{test}}\right) = \frac{\left(A^t w_{\text{target}}\right)^t W^{-1} \left(A^t w_{\text{test}}\right)}{\sqrt{\left(A^t w_{\text{target}}\right)^t W^{-1} \left(A^t w_{\text{target}}\right)} \cdot \sqrt{\left(A^t w_{\text{test}}\right)^t W^{-1} \left(A^t w_{\text{test}}\right)}} \tag{8}$$

3. What score normalization is doing in cosine similarity scoring?

In this section, we analyze how score normalization techniques are reflected in the ivector cosine similarity to derive a new extended cosine similarity that does not require explicit score normalization. We first define $w' = \frac{B^t A^t w}{\left\|B^t A^t w\right\|}$, where A is the

LDA projection matrix and $BB^t = W^{-1}$ being the Cholesky decomposition of the inverse WCCN matrix. The cosine similarity is then simplified to a dot product between a normalized target ivector w'_{target} and a test ivector w'_{test},

$$\text{score}\left(w'_{\text{target}}, w'_{\text{test}}\right) = \left(w'_{\text{target}}\right)^t \left(w'_{\text{test}}\right) \tag{9}$$

The z-normalized score is

$$s\text{znorm} = \frac{\text{score}\left(w'_{\text{target}}, w'_{\text{test}}\right) - \mu\text{z\_norm}}{\sigma\text{z\_norm}} \tag{10}$$

where μz_norm and σz_norm are the score normalization parameters for z-norm relative to the target. These parameters are computed over a set of utterances from speakers different than the target speaker (i.e., impostor or non-target speakers),

$$\begin{aligned}\mu_{\text{target}} &= \frac{\left(w'_{\text{target}}\right)^t \left(w'_{\text{z\_impl1}}\right) + \left(w'_{\text{target}}\right)^t \left(w'_{\text{z\_impl2}}\right) + \cdots + \left(w'_{\text{target}}\right)^t \left(w'_{\text{z\_implN}}\right)}{N} \\ &= \left(w'_{\text{target}}\right)^t \overline{w'_{\text{z\_imp}}}\end{aligned} \tag{11}$$

Where $\overline{w'_{\text{z\_imp}}} = \frac{\left(w'_{\text{z\_impl1}}\right)+\left(w'_{\text{z\_impl2}}\right)+\ldots+\left(w'_{\text{z\_implN}}\right)}{N}$ corresponds to the mean of z-norm impostor ivectors. As we do with LDA, we estimate the impostor ivector mean rather than assume it to be equal to zero.

The second z-norm parameter can be obtained by

$$\begin{aligned}\sigma^2 &= E_i\Big[\left(\left(w'_{\text{target}}\right)^t \left(w'_{\text{z\_imp}_i}\right) - \mu_{\text{target}}\right) \\ &\qquad * \left(\left(w'_{\text{target}}\right)^t \left(w'_{\text{z\_imp}_i}\right) - \mu_{\text{target}}\right)^t\Big] \\ &= E_i\Big[\left(\left(w'_{\text{target}}\right)^t \left(w'_{\text{z\_imp}_i}\right) - \left(\left(w'_{\text{target}}\right)^t \left(\overline{w'_{\text{z\_imp}}}\right)\right)\right) \\ &\qquad * \left(\left(w'_{\text{target}}\right)^t \left(w'_{\text{z\_imp}_i}\right) - \left(\left(w'_{\text{target}}\right)^t \left(\overline{w'_{\text{z\_imp}}}\right)\right)\right)^t\Big] \\ &= \left(w'_{\text{target}}\right)^t \Sigma_{\text{z\_imp}} \left(w'_{\text{target}}\right)\end{aligned} \tag{12}$$

Where $\Sigma_{\text{z\_imp}} = E_i\left[\left(\left(w'_{\text{z\_imp}_i}\right) - \overline{w'_{\text{z\_imp}}}\right)\left(\left(w'_{\text{z\_imp}_i}\right) - \overline{w'_{\text{z\_imp}}}\right)^t\right]$ is the covariance matrix of the set of z-norm impostor ivectors. Thus equation (10) can be rewritten as

$$\begin{aligned}s\text{znorm} &= \frac{\text{score}\left(w'_{\text{target}}, w'_{\text{test}}\right) - \left(w'_{\text{target}}\right)^t \overline{w'_{\text{z\_imp}}}}{\sqrt{\left(w'_{\text{target}}\right)^t \Sigma_{\text{z\_imp}} \left(w'_{\text{target}}\right)}} \\ &= \frac{\left(w'_{\text{target}}\right)^t \left(w'_{\text{test}} - \overline{w'_{\text{z\_imp}}}\right)}{\sqrt{\left(w'_{\text{target}}\right)^t \Sigma_{\text{z\_imp}} \left(w'_{\text{target}}\right)}}\end{aligned} \tag{13}$$

If we suppose that $\Sigma_{\text{z\_imp}}$ is a diagonal matrix, the last equation can be simplified to

$$s\text{znorm} = \frac{\left(w'_{\text{target}}\right)^t \left(w'_{\text{test}} - \overline{w'_{\text{z\_imp}}}\right)}{\left\|C_{\text{z\_imp}} w'_{\text{target}}\right\|} \tag{14}$$

Where $C_{\text{z\_imp}}$ is diagonal matrix which contains the square root of diagonal z-norm impostor's covariance matrix, $\Sigma_{\text{z\_imp}}$. In summary, z-norm score normalization is shifting the test ivector w'_{test} by the mean of the z-norm impostor projected ivectors and scaling by the length of the target ivector based on z-norm impostor covariance.

In a similar manner, we can obtain the score normalization parameters for t-norm and the final normalized score can be expressed by

$$s_{\text{tnorm}} = \frac{\left(w'_{\text{target}} - \overline{w'_{\text{t\_imp}}}\right)^t \left(w'_{\text{test}}\right)}{\left\|C_{\text{t\_imp}} w'_{\text{test}}\right\|} \tag{15}$$

Table 1: Corpora used to estimate the UBM, total variability matrix (T), LDA and WCCN

	UBM	T	LDA	WCCN
Switchbord II, Phase 2 and 3	X	X	X	
Switchboard II Cellular, Part 1 and 2	X	X	X	
Fisher English database Part 1 and 2		X		
NIST 2004 SRE	X	X	X	X
NIST 2005 SRE	X	X	X	X

Where $C_{\text{t\_imp}}$ is diagonal matrix which contains the square root of diagonal t-norm impostor's covariance matrix, $\Sigma_{\text{t\_imp}}$. Comparing the final normalized scores, we see that z- and t-norm in the ivector space are just duals of each other merely switching the roles of target and test ivectors.

4. New scoring

In this section, we propose a new cosine similarity scoring equation that combines the effects of z- and t-norm score normalization. This new scoring, given below in Equation 16, does not require test-time score normalization compared to the classical cosine similarity scoring proposed in our previous works [3, 1].

$$\text{score}\left(w'_{\text{target}}, w'_{\text{test}}\right) = \frac{\left(w'_{\text{target}} - \overline{w'_{\text{imp}}}\right)^t \left(w'_{\text{test}} - \overline{w'_{\text{imp}}}\right)}{\left\|C_{\text{imp}} \cdot w'_{\text{target}}\right\| \left\|C_{\text{imp}} \cdot w'_{\text{test}}\right\|} \tag{16}$$

where $\overline{w'_{\text{imp}}}$ is the mean of the impostor ivectors. C_{imp} is a diagonal matrix that contains the square root of the diagonal covariance matrix of the impostor ivectors.

5. Experiment

5.1. Experimental set-up

Our experiments operate on cepstral features, extracted using a 25 ms Hamming window. 19 mel frequency cepstral coefficients together with log energy are calculated every 10 ms. This 20-dimensional feature vector was subjected to feature warping [8] using a 3 s sliding window. Delta and double delta coefficients were then calculated using a 5 frame window to produce 60-dimensional feature vectors. We used gender dependent universal background models containing 2048 Gaussians. Table 1 summarizes all corpora are used to estimate the UBM, total variability matrix, LDA and WCCN. The choice of these corpora is described in [3, 1]

The baseline system uses original cosine scoring with zt-norm score normalization. We used 250 t-norm impostor models taken from NIST 2005 SRE data and 1200 impostor models taken from Switchboard II, Phases 2 and 3; Switchboard Cellular, Parts 1 and 2; and NIST 2004 SRE data. In the new scoring system, we used all previous impostors in the baseline system to estimate the impostor mean and covariance matrix.

5.2. Results

All our experiments were carried out on the female telephone data part of the 1 conversation train/1 conversation test (nominally 2.5 min in train and test, 1conv-1conv core) and 10 second train/10 second test (10sec-10sec) conditions of the 2008 SRE

Table 2: Comparison of results between original cosine similarity scoring and the new scoring that incorporates score normalization. The results are on the female portion of the 1conv-1conv 2008 SRE core telephone condition.

	English trials		All trials	
	EER	DCF	EER	DCF
Origial cosine similarity scoring with zt-norm	**2.90%**	**0.0124**	5.76%	0.0322
New cosine similarity	3.41%	0.0127	**5.21%**	**0.0248**

Table 3: Comparison results between original cosine similarity scoring and the new scoring that incorporates score normalization. The results are on the female portion of the 10sec-10sec 2008 SRE telephone condition.

	English trials		All trials	
	EER	DCF	EER	DCF
Origial cosine similarity scoring with zt-norm	12.10%	0.0577	16.59%	0.0725
New cosine similarity	**11.42%**	**0.0573**	**15.83%**	**0.0673**

dataset. We compared the results obtained from the original cosine similarity scoring with zt-norm score normalization and the new cosine scoring that incorporates score normalization. The results are reported in Tables 2 and 3.

The results obtained with these two systems show that the new scoring generally gives the best results especially for all trials pooling. We achieved a DCF of 0.0248 with the new scoring compared to 0.0322 obtained with original cosine similarity and zt-norm score normalization. However, the original cosine similarity obtained the best EER for English trials of the core condition. In the next section, we will describe how we use the new scoring with a simple speaker unsupervised adaptation algorithm.

5.3. Speaker unsupervised adaptation:

In [4], we propose a new unsupervised speaker adaptation algorithm based on the cosine similarity computed between ivectors. The two challenging problems in the unsupervised adaptation problem are (1) updating both the score normalization parameters and (2) setting the decision threshold to allow adaptation. This new adaptation algorithm propose an easier way to update the score normalization parameters and the optimal decision threshold corresponds to one which minimize the DCF on the development dataset. Using this new scoring, the unsupervised speaker adaptation algorithm becomes less complex because there is no score normalization parameters updating compared to the previous cosine similarity scoring. When we have a target and test ivectors, the cosine similarity is computed and compared to the decision threshold. If it is greater than the threshold, the test ivector is considered as a target ivector and is kept with the original target ivector to represent the target speaker. When we have another test ivector to compare with the target speaker; we compute the cosine similarity to each target

Table 4: Comparison results between the original cosine similarity scoring and the new scoring using the unsupervised speaker adaptation algorithm. The results are on the female portion of the 1conv-1conv 2008 SRE core telephone condition.

	English trials		All trials	
	EER	DCF	EER	DCF
Origial cosine similarity scoring with zt-norm	**2.90%**	0.0124	5.76%	0.0322
New cosine	3.41%	0.0127	5.21%	0.0248
Similarity Without adaptation				
New cosine similarity With adaptation	3.17%	**0.0107**	**4.83%**	**0.0229**

Table 5: Comparison results between original cosine similarity scoring and the new scoring using the unsupervised speaker adaptation algorithm. The results are on the female portion of the 10sec-10sec 2008 SRE telephone condition.

	English trials		All trials	
	EER	DCF	EER	DCF
Origial cosine similarity scoring with zt-norm	12.10%	0.0577	16.59%	0.0725
New cosine similarity Without adatation	11.42%	0.0573	15.83%	0.0673
New cosine similarity With adaptation	**10.68%**	**0.0560**	**15.42%**	**0.0660**

speaker ivectors and the final score is equal to the mean of the two scores. This new score is compared to the decision threshold and, if greater, the test ivector is added to the target list. The process continues for all test utterances. The results of the unsupervised speaker algorithm based on the new scoring are reported in Table 4 and 5. Note that the results are obtained on the NIST 2008 SRE, but the decision threshold is estimated on a development dataset from the NIST 2006 SRE. All unsupervised speaker adaptation experiments are carried out with the respect to the NIST 2008 SRE unsupervised adaptation protocol.

The results given in Tables 4 and 5 show an improvement by using the unsupervised adaptation algorithm, especially for all trials pooling. The EER for the 1conv-1conv condition decreases from 5.21% (see Table 2) to 4.83% (see Table 4). The same behavior was noted for the 10sec-10sec condition. We also achieved and Minimum DCF of 0.0107 for female part of the English core condition. Figure 1 shows the both detcurves with and without speaker unsupervised adaptation on the English female part of the NIST 2008 SRE.

6. conclusion

In speaker recognition systems using JFA, score normalization is critical for good performance. In this paper, we proposed an extension of the cosine similarity to remove explicit score normalization from the decision process. This new scoring sim-

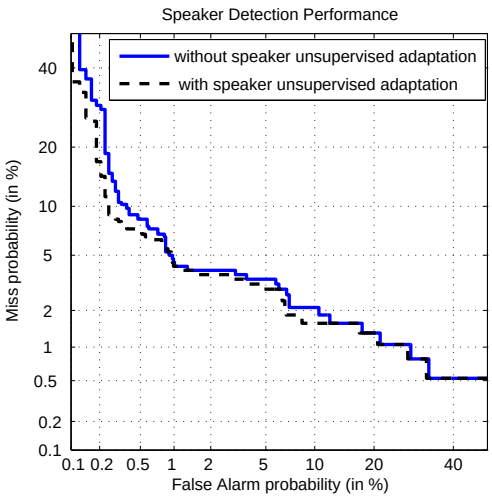

Figure 1: *Detcurves on the English female part of the NIST 2008 Speaker recognition evaluation.*

ulates directly zt-norm. The mean and covariance matrix computed on a set of impostor ivectors are used to normalize the cosine similarity. This new scoring simplifies the speaker unsupervised adaptation algorithm already proposed for cosine similarity scoring. We achieved a MinDCF of 0.0107 on female English part of the NIST 2008 speaker recognition evaluation.

7. References

[1] Najim Dehak, *Discriminative and Generative Approaches for Long- and Short-Term Speaker Characteristics Modeling: Application to Speaker Verification*, Ph.D. thesis, École de Technologie Supérieure, Montreal, 2009.

[2] P. Kenny, P. Ouellet, N. Dehak, V. Gupta, and P. Dumouchel, "A Study of Interspeaker Variability in Speaker Verification," *IEEE Transaction on Audio, Speech and Language*, vol. 16, no. 5, pp. 980–988, july 2008.

[3] Najim Dehak, Patrick Kenny, Rda Dehak, Pierre Ouellet, and Pierre Dumouchel, "Front end Factor Analysis for Speaker Verification," submitted to IEEE Transactions on Audio, Speech and Language Processing, 2010.

[4] Stephen Hin-Chung Shum, Najim Dehak, Reda Dehak, and Jim Glass, "Unsupervised Speaker Adaptation based on the Cosine Similarity for Text-Independent Speaker Verification," Submitted to ODYSSEY, 2010.

[5] O. Glembek, L. Burget, N. Brummer, and P. Kenny, "Comparaison of Scoring Methods Used in Speaker Recognition with Joint Factor Analysis," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Taipei, Taiwan, april 2009.

[6] Najim Dehak, Rda Dehak, Patrick Kenny, Niko Brummer, Pierre Ouellet, and Pierre Dumouchel, "Support Vector Machines versus Fast Scoring in the Low-Dimensional Total Variability Space for Speaker Verification," in *INTERSPEECH*, Brighton, UK, september 2009.

[7] A. Hatch, S. Kajarekar, and A. Stolcke, "Within-Class Covariance Normalization for SVM-Based Speaker Recognition," in *International Conference on Spoken Language Processing*, Pittsburgh, PA, USA, September 2006.

[8] J. Pelecanos and S. Sridharan, "Feature Warping for Robust Speaker Verification," in *IEEE Odyssey: The Speaker and Language Recognition Workshop*, Crete, Greece, 2001, pp. 213–218.