Modified-prior i-Vector Estimation for Language Identification of Short Duration Utterances

Ruchir Travadi, Maarten Van Segbroeck, and Shrikanth Narayanan

Signal Analysis and Interpretation Lab
University of Southern California, Los Angeles, CA – 90089
travadi@usc.edu, {maarten, shri}@sipi.usc.edu

Abstract

In this paper, we address the problem of Language Identification (LID) on short duration segments. Current state-of-the-art LID systems typically employ total variability i-Vector modeling for obtaining fixed length representation of utterances. However, when the utterances are short, only a small amount of data is available, and the estimated i-Vector representation will consequently exhibit significant variability, making the identification problem challenging. In this paper, we propose novel techniques to modify the standard normal prior distribution of the i-Vectors, to obtain a more discriminative i-Vector extraction given the small amount of available utterance data. Improved performance was observed by using the proposed i-Vector estimation techniques on short segments of the DARPA RATS corpora, with lengths as small as 3 seconds.

Index Terms: Language identification, i-Vector, short duration segments, RATS

1. Introduction

Language Identification (LID) refers to the problem of automatically identifying the source language from a given speech utterance. Over the years, several techniques have been proposed for this problem, ranging from those which leverage phonotactic information, such as PRLM and PPRLM [1], to more recent acoustic modeling approaches, which focus on spectral characteristics of utterances. Acoustic modeling approaches typically reduce an utterance to a representation which captures the variability of its Gaussian Mixture Model (GMM) supervectors. Joint Factor Analysis (JFA) [2], which maps the GMM supervectors into separate source and channel variability factors, is one of the popular approaches in this domain.

Total variability i-Vector modeling [3, 4], an approach which evolved from JFA, has gained popularity as an elegant framework to obtain a fixed dimensional representation of variable length speech utterances. It hypothesizes that both source and channel variabilities of the utterances lie in a single low dimensional subspace, known as the total variability or i-Vector space. Every utterance is represented as an i-Vector, which is the posterior expectation of its corresponding GMM supervector representation projected onto the total variability space. Compensation methods such as Linear Discriminant Analysis (LDA), Within Class Covariance Normalization (WCCN) [5] and Nuisance Attribute Projection (NAP) [6] have been suggested to tackle the problem of inter-session variability in the i-Vector space. Classification is then performed on extracted i-Vectors to find target classes using classification tools such as PLDA [7, 8], SVM [9] or Neural Networks [10].

When the utterances used for classification are of significantly long duration, the i-Vectors exhibit low intra-class variability and result in compact clusters which are fairly separable. However, one of the major challenges faced by these systems is the significant performance degradation on short duration segments. This challenging task of accurate classification on short duration utterances using i-Vector framework has previously been addressed in [11–14], and in [15], where a novel UBM fusion approach is adopted. In this paper, we propose to tackle this problem by constraining the prior distribution of the i-Vector estimates to restrict their variability.

The remainder of this paper is organized as follows : Section 2 describes the baseline i-Vector extraction scheme. In section 3, we motivate the use of a different prior, and derive the expressions for i-Vector estimation under modified prior assumptions. Section 4 describes the proposed techniques for parameter selection of modified prior distribution. The experimental setup has been laid out in section 5. The obtained results are summarized in section 6. Conclusion and discussion of future avenues are given in section 7.

2. Baseline system

2.1. Total Variability i-Vector Modeling

Suppose we are given a UBM consisting of C components and parameter set $\mathbf{\Lambda} = \{\boldsymbol{\lambda_1}, \boldsymbol{\lambda_2}, ..., \boldsymbol{\lambda_C}\}$, where each mixture component c is characterized by $\boldsymbol{\lambda_c} = \{\rho_c, \boldsymbol{\mu_c}, \boldsymbol{\Sigma_c}\}$. Then, for an utterance $\boldsymbol{y_j}$ with feature sequence $[\boldsymbol{y_{j,1}}, \boldsymbol{y_{j,2}}, ..., \boldsymbol{y_{j,L}}]$, the zeroth and first order centered Baum Welch statistics, N_c^j and F_c^j respectively, for each component c are obtained as follows :

$$N_c^j = \sum_{t=1}^{L} P\left(c|\boldsymbol{y_{j,t}}, \boldsymbol{\lambda_c}\right) \quad (1)$$

$$F_c^j = \frac{1}{N_c^j} \sum_{t=1}^{L} P\left(c|\boldsymbol{y_{j,t}}, \boldsymbol{\lambda_c}\right)\left(\boldsymbol{y_{j,t}} - \boldsymbol{\mu_c}\right) \quad (2)$$

After concatenating the statistics for each UBM component, we obtain the diagonal matrix of occupancy counts $\mathbf{N}$, and the supervector $\mathbf{F}$. In total variability i-Vector modeling, it is hypothesized that the $\mathbf{F} = \mathbf{Tx}$, where $\mathbf{T}$ is the total variability matrix of low rank and $\mathbf{x}$ is known as the i-Vector. This can be re-interpreted as a classic factor analysis based generative modeling problem, where it is assumed that the prior distribution of i-Vectors and the distribution of $\mathbf{F}$ conditioned on the i-Vector are Gaussian [16] :

$$P(\boldsymbol{x}) = \mathcal{N}(\mathbf{0}, \mathbf{I}), \qquad P\left(\boldsymbol{F}|\boldsymbol{x}\right) = \mathcal{N}\left(\mathbf{Tx}, \boldsymbol{N}^{-1}\boldsymbol{\Sigma}\right) \quad (3)$$

Given prior assumptions of (3), the MAP estimate of $\boldsymbol{x}$ given $\boldsymbol{F}$

is given as

$$E\left[\boldsymbol{x}|\boldsymbol{F}\right] = \left(\mathbf{I} + \mathbf{T}^t \boldsymbol{\Sigma}^{-1} \boldsymbol{NT}\right)^{-1} \mathbf{T}^t \boldsymbol{\Sigma}^{-1} \boldsymbol{NF} \qquad (4)$$

The estimate given by (4) represents the mean of the posterior distribution of $\mathbf{x}$ given $\mathbf{F}$, and is adopted as the i-Vector representation of the utterance.

2.2. Simplified i-Vector Modeling

The i-Vector extraction procedure corresponding to (4) is computationally very expensive. For a UBM with C components, feature dimension D and i-Vector dimension K, the computational complexity is $O(K^3 + K^2C + KCD)$ [17]. Li et al [16] addressed this problem by suggesting a simplified i-Vector extraction procedure, which achieves a significant speedup in i-Vector extraction, reducing the complexity to $O(K^3 + KCD)$, at the cost of only a negligible degradation in accuracy. In this framework, for an utterance with n frames, the $\mathbf{F}$ vectors are first normalized by occupancy counts :

$$\boldsymbol{F_c^j} \leftarrow \sqrt{m_c^j} \boldsymbol{F_c^j}, \qquad \text{where } m_c^j = \frac{N_c^j}{\sum_{c=1}^C N_c^j} \qquad (5)$$

Then, the covariance of the conditional distribution of $\mathbf{F}$ is simplified as $n^{-1}\boldsymbol{\Sigma}$. In the simplified framework, i-Vectors are now estimated as :

$$E\left[\boldsymbol{x}|\boldsymbol{F}\right] = \left(\mathbf{I} + \mathbf{T}^t \boldsymbol{\Sigma}^{-1} n\mathbf{T}\right)^{-1} \mathbf{T}^t \boldsymbol{\Sigma}^{-1} n\boldsymbol{F} \qquad (6)$$

We use the i-Vector extraction in (6) as our baseline estimate.

3. I-Vector Extraction With Prior Modification

3.1. Motivation

The implicit assumption on i-Vectors estimated from the data is that utterances belonging to the same class would map to similar i-Vectors. Therefore, it is ideally expected that the collection of i-Vectors in data would form a set of compact, separable clusters, each of which corresponds to a particular class. However, the i-Vector extraction procedure fails to make this assumption explicit. The hypothesized standard normal prior on i-Vectors only serves as a check on the magnitude of the obtained i-Vectors. It does not promote any form of clustering for i-Vectors to be estimated from the data.

As the utterance length increases, typically the within-class variability in estimated sufficient statistics reduces, and the obtained i-Vectors tend to form clusters. However, this is not the case for i-Vectors estimated from short duration data, because of significant within-class variability in estimated sufficient statistics. In such cases, it might be beneficial to make the assumption about presence of distinct clusters in the data explicit through the prior itself.

Suppose there are M classes in the data, with corresponding prior probabilities $\{P_C(i)\}_{i=1}^M$. We hypothesize that that i-Vectors from class i form a cluster with mean $\mu_{\mathbf{i}}$ and covariance $\mathbf{C}_i$. Then, the hypothesized prior distribution on the i-Vectors without any label information is a GMM prior $\sum_{i=1}^M P_C(i)\mathcal{N}(\mu_i, \mathbf{C}_i)$. If the label corresponding to an i-Vector estimate is known to be i, the hypothesized prior distribution on the i-Vector is normal distribution $\mathcal{N}(\mu_i, \mathbf{C}_i)$. Therefore, the assumption is that there are as many Gaussian mixtures in the prior distribution as there are classes in the data. While

the derivations that follow extend to arbitrary number of mixtures, we use same number of mixtures as there are classes, for our analysis in this paper.

In 3.2 and 3.3, we derive expressions for estimating i-Vectors under these prior assumptions. Then, various techniques for selecting appropriate parameters for the prior distribution are proposed in section 4.

3.2. General Gaussian prior

Under assumption of Gaussian prior with arbitrary mean μ and an arbitrary covariance matrix $\mathbf{C}$:

$$P(\mathbf{x}) = \mathcal{N}(\mu, \mathbf{C}), \quad P(\mathbf{F}|\mathbf{x}) = \mathcal{N}(\mathbf{Tx}, n^{-1}\boldsymbol{\Sigma}) \qquad (7)$$

Then, the posterior distribution of $\mathbf{x}$ is given as :

$$\begin{aligned}
P\left(\mathbf{x}|\mathbf{F}\right) &= P\left(\mathbf{F}|\mathbf{x}\right) P\left(\mathbf{x}\right) \\
&\propto \exp\left[-\frac{1}{2}\left(\mathbf{F}-\mathbf{Tx}\right)^t \boldsymbol{\Sigma}^{-1} n\left(\mathbf{F}-\mathbf{Tx}\right)\right. \\
&\qquad \left. -\frac{1}{2}\left(\mathbf{x}-\mu\right)^t \mathbf{C}^{-1}\left(\mathbf{x}-\mu\right)\right] \\
&\propto \exp\left[\mathbf{x}^t \mathbf{T}^t \boldsymbol{\Sigma}^{-1} n\mathbf{F} - \frac{1}{2}\mathbf{x}^t \mathbf{T}^t \boldsymbol{\Sigma}^{-1} n\mathbf{Tx}\right. \\
&\qquad \left. -\frac{1}{2}\mathbf{x}^t \mathbf{C}^{-1}\mathbf{x} + \mathbf{x}^t \mathbf{C}^{-1}\mu\right] \\
&\propto \exp\left[\mathbf{x}^t \left(\mathbf{T}^t \boldsymbol{\Sigma}^{-1} n\mathbf{F} + \mathbf{C}^{-1}\mu\right)\right. \\
&\qquad \left. -\frac{1}{2}\mathbf{x}^t \left(\mathbf{T}^t \boldsymbol{\Sigma}^{-1} n\mathbf{T} + \mathbf{C}^{-1}\right)\mathbf{x}\right] \\
&\propto \exp\left[-\frac{1}{2}\left(\mathbf{x}-\boldsymbol{l}^{-1}\mathbf{b}\right)^t \boldsymbol{l}\left(\mathbf{x}-\boldsymbol{l}^{-1}\mathbf{b}\right)\right] \qquad (8)
\end{aligned}$$

where

$$\mathbf{b} = \mathbf{T}^t \boldsymbol{\Sigma}^{-1} n\mathbf{F} + \mathbf{C}^{-1}\mu, \quad \boldsymbol{l} = \mathbf{C}^{-1} + \mathbf{T}^t \boldsymbol{\Sigma}^{-1} n\mathbf{T} \qquad (9)$$

Therefore, $P(\mathbf{x}|\mathbf{F}) = \mathcal{N}(\boldsymbol{l}^{-1}\mathbf{b}, \boldsymbol{l}^{-1})$, and hence i-Vectors are estimated as :

$$E\left[\mathbf{x}|\mathbf{F}\right] = \boldsymbol{l}^{-1}\mathbf{b} \qquad (10)$$

3.3. Gaussian Mixture Prior

Under Gaussian Mixture prior assumption with class probabilities $P_C(i)$, means μ_i, covariances $\mathbf{C}_i$:

$$P(\mathbf{x}) = \sum_{i=1}^M P_C(i)\mathcal{N}(\mu_i, \mathbf{C}_i) \qquad (11)$$

Using a similar approach to (9), we can derive that the posterior distribution is given by

$$P(\mathbf{x}|\mathbf{F}) = \sum_{i=1}^M P_C(i)\mathcal{N}(\boldsymbol{l_i}^{-1}\mathbf{b}_i, \boldsymbol{l_i}^{-1}) \qquad (12)$$

where,

$$\mathbf{b}_i = \mathbf{T}^t \boldsymbol{\Sigma}^{-1} n\mathbf{F} + \mathbf{C}_i^{-1}\mu_i, \quad \boldsymbol{l_i} = \mathbf{C}_i^{-1} + \mathbf{T}^t \boldsymbol{\Sigma}^{-1} n\mathbf{T} \qquad (13)$$

The estimated i-Vector x is then given as :

$$E\left[\mathbf{x}|\mathbf{F}\right] = \sum_{i=1}^M P_C(i)\,\boldsymbol{l}_i^{-1}\mathbf{b}_i \qquad (14)$$

3.4. Prior reweighting

The i-Vector estimates given by (10) and (14) are combinations of terms given by observed statistics and prior parameters. Additional flexibility can be gained on balancing the impact of the imposed prior against the observed data by introducing a parameter λ, which controls the relative weights of each term in the i-Vector estimate. The i-Vector estimates for general Gaussian prior using parameter λ are obtained as :

$$E\left[\mathbf{x}|\mathbf{F}\right] = (\mathbf{C}^{-1} + \lambda\mathbf{T}^t\mathbf{\Sigma}^{-1}n\mathbf{T})^{-1}(\lambda\mathbf{T}^t\mathbf{\Sigma}^{-1}n\mathbf{F} + \mathbf{C}^{-1}\mu) \tag{15}$$

For GMM prior, the i-Vectors are obtained as :

$$E\left[\mathbf{x}|\mathbf{F}\right] = \sum_{i=1}^{M} P_C(i)(\mathbf{C}_i^{-1} + \lambda\mathbf{T}^t\mathbf{\Sigma}^{-1}n\mathbf{T})^{-1}$$
$$(\lambda\mathbf{T}^t\mathbf{\Sigma}^{-1}n\mathbf{F} + \mathbf{C}_i^{-1}\mu_i) \tag{16}$$

Adjusting λ allows us to control the degree of penalization on deviations from the prior assumption. A value of $\lambda < 1$ would emphasize the prior, penalizing the deviations from mixture means more severely. A value of $\lambda > 1$ would emphasize the data term higher, relaxing the penalty imposed by the prior.

The results reported in Section 6 correspond to the choice of λ that gave best results. Typically, a choice of λ in the range of 0.1 to 0.4 was found to be optimal, indicating that emphasizing the prior was useful.

4. Prior Parameter Selection

4.1. GMM prior parameters derived from long duration segments

We hypothesize that i-Vectors estimated from long duration segments exhibit low within-class variability, and i-Vectors from these utterances can be used to estimate mixture means and covariances μ_i and $\mathbf{C}_i$ for GMM-prior based estimation.

Therefore, for this approach, we first obtain i-Vectors for the long duration training examples according to standard normal prior. Then, we evaluate mean and covariance estimates μ_i and $\mathbf{C}_i$ for each class i, $i \in \{1, 2, ..., M\}$. These parameters, along with class priors $P_C(i)$ are then used for estimating i-Vectors, using equation (16). This approach is motivated by the idea that the i-Vector estimate for an utterance is unlikely to belong to a region which does not lie close to one of the clusters observed in training data from long duration segments. The imposed prior acts as a filter to reduce the density of posterior estimate in regions that are unlikely. For reporting the results in Section 6, we refer to this approach as the GMM i-Vector approach.

4.2. Classifier output based re-estimation

4.2.1. Gaussian Mixture Re-estimation

Suppose we have performed one stage of classification on a given test utterance already, and the probabilities of the utterance belonging to a particular class are available from the classifier output. Then, we can use can substitute mixture weights in GMM prior by the class probabilities from classifier output, as opposed to using class priors, to estimate the i-Vector once again for another stage of classification. This GMM prior not only penalizes the estimate for being far away from clusters observed in long duration training data, it also attempts to nullify the influence of mixtures corresponding to classes that

were deemed unlikely for the utterance by the classifier. For reporting results in Section 6, we refer to this approach as GMM Re-estimation approach.

4.2.2. Score combination of individual class priors

A closer look at (16) reveals that i-Vector estimate under GMM prior is just a weighted linear combination of estimates under assumption of single Gaussian prior corresponding to individual mixture components, weighted by $P_C(i)$. Therefore, the procedure suggested in Section 4.2.1 amounts to re-estimating the i-Vector under assumption of a single Gaussian corresponding to each class $i \in \{1, 2, ..., M\}$, and then combining the estimates using available class probabilities as weights.

This motivates the idea that the combination can also be performed at score level, rather than the i-Vector estimate level. Therefore, in this approach, for each utterance, we obtain M i-Vector estimates, $\{\mathbf{x}_i\}_{i=1}^{M}$, where $\mathbf{x}_i$ is obtained using a single Gaussian prior corresponding to class i, with mean μ_i and covariance $\mathbf{C}_i$, according to (15). Then, we perform classification separately on each of the obtained estimates $\mathbf{x}_i$, to obtain probabilities $s_{ij} = P(\mathbf{x}_i \in classj)$.

Let $P_C(i)$ denote the class probabilities available from first-pass classification. Then, we obtain updated score as the weighted sum of all the scores $\{s_{ij}\}_{i=1}^{M}$ as follows :

$$\hat{P}_C(i) = \sum_{i=1}^{M} P_C(i)s_{ij} \tag{17}$$

For reporting results in Section 6, we refer to this approach as Score Re-estimation approach.

5. Experimental Setup

We used the DARPA Robust Automatic Transcription of Speech (RATS) data corpus [18] for our experiments, which consists of noisy audio recordings from six classes : five target languages and a class corresponding to 10 non-target languages. For each frame, we obtained 22-dimensional MFCC vectors, concatenated with delta coefficients, resulting in a 44-dimensional feature vector. A UBM of 2048 components was trained on all available speech data, and an i-Vector dimension of 400 was used. WCCN was applied on the i-Vector space for intersession variability compensation. The i-Vectors were subsequently classified by training an SVM with a fifth order polynomial kernel. The training data consisted of collection of segments of length 30s, 10s and 3s. The test set DEV2 consisted of 1335 samples of duration 3s. In the next section, the following measures are reported on samples of the DEV2 set : EER, DCF, P_{miss} for 10% False Alarm rate (denoted by P_{miss}^{10}), and classification accuracy.

6. Results

For our experiments, we used the simplified i-Vector extraction scheme [16] as our baseline. First, we performed three separate experiments by using 30s, 10s and 3s data respectively for obtaining $\mathbf{T}$ and $\mathbf{\Sigma}$ parameters through EM. We then used the obtained estimates to extract i-Vectors on 3s training utterances to train the SVM. Experimental results (not reported in this paper) showed that duration matched training for SVM achieved best results. The results obtained using duration matched SVM training have been reported in Baseline row of Table 1. The column D_T in Table 1 indicates the duration length of training samples (in seconds) used for EM procedure.

It is apparent from the table that using 30s segments for EM training gave the best results. This is intuitive since the EM algorithm used for estimation of subspace parameters $\mathbf{T}$ and $\mathbf{\Sigma}$ also depends on intermediate evaluation of i-Vectors, which are highly variable for short duration data. Once we have the best subspace estimate trained on longer duration data, the classifier is best trained on short duration segments, as mismatch in distributions between training and test sets affects the SVM adversely.

Having established the best baseline result, we then applied the various methods suggested in section 4 for i-Vector extraction. For the GMM i-Vector approach (Section 4.1), we obtained μ_i and $\mathbf{C}_i$ estimates from 30s training set, and the class priors $P_C(i)$ were obtained as proportion of training data belonging to class i. For GMM Re-estimation (Section 4.2.1) and Score Re-estimation (Section 4.2.2), we used SVM outputs from the best baseline system corresponding to 30s EM training as class probabilities, and performed the suggested re-estimation and scoring. The results have been summarized in Table 1 below.

System	D_T	EER	DCF	P_{miss}^{10}	Accuracy
	3	17.89	17.76	25.85	65.47
Baseline	10	16.71	16.45	24.41	68.31
	30	15.40	15.21	22.19	69.74
GMM i-Vector (4.1)	30	15.27	14.98	20.76	69.44
GMM Re-estimation (4.2.1)	30	16.32	15.82	22.32	70.11
Score Re-estimation (4.2.2)	30	15.14	15.07	21.28	69.96

Table 1: *Experimental Results*

It can be seen that a noticeable performance gain was observed in P_{miss}^{10} using the GMM i-Vector estimation technique, along with minor gains in EER and DCF. However, the classification accuracy decreased slightly. One of the possible factors that limits the performance gained using this approach is related to the extraction of point-estimate representation of i-Vectors from the posterior distribution. Although the modification in prior attempts to concentrate the density of posterior distribution near the clusters observed in 30s training data, the SVM classification step requires point estimates, which are obtained by evaluating the mean of the posterior distribution. However, the mean of a GMM distribution is not guranted to lie in a region of high density, and therefore i-Vector estimates obtained using this prior are not necessarily confined to clusters observed in the training data.

While the GMM Re-estimation approach for second classification pass resulted in higher classification accuracy, it was only achieved at the cost of significant degradation in EER, DCF and P_{miss}^{10} values. The intuition behind the observed degradation is that if the classifier assigns very high probability for the wrong class in first pass of classification, then the re-estimated i-Vector gets heavily biased towards the cluster mean of that class. Hence, the confidence in the wrong label increases further, which affects EER, DCF and P_{miss}^{10} negatively, but does not affect classification accuracy since the obtained label value does not change after re-estimation. However, if the true class gets assigned a reasonably high probability in the first pass, then the re-estimation process might help guiding the updated estimate towards the cluster mean of the true class by removing the contribution of unlikely classes from the estimate.

The Score Re-estimation technique resulted in moderate gains in EER, DCF, P_{miss}^{10} as well as classification accuracy, showing that score level combination was more helpful than i-Vector estimate level combination.

7. Conclusion and Future Direction

In this paper, we have described novel approaches to tackle the problem of i-Vector variability of short duration segments by modifying prior distributions of the i-Vectors. We derived an expression for i-Vector estimation by assuming the prior to be Gaussian distributed with arbitrary mean and covariance, or modeled by a GMM. We used the class mean and covariance estimates from longer duration segments to obtain GMM prior for shorter duration segments, using either class prior probabilities or available class probability estimates of a classifier as mixture weights.

Currently, we use standard normal priors within the EM estimation scheme for obtaining total variability subspace parameters $\mathbf{T}$ and $\mathbf{\Sigma}$, and use modified priors only for i-Vector extraction with pre-calculated $\mathbf{T}$ and $\mathbf{\Sigma}$ estimates. However, the baseline results in Table 1 show that better estimates of $\mathbf{T}$ and $\mathbf{\Sigma}$ obtained using longer duration segments had a remarkable impact on the system performance. Therefore, as future work, we plan to extend the use of prior modificaion in EM training.

8. Acknowledgements

9. References

[1] M. A. Zissman, "Language identification using phoneme recognition and phonotactic language modeling," in *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on*, vol. 5. IEEE, 1995, pp. 3503–3506.

[2] P. Kenny, G. Boulianne, and P. Dumouchel, "Eigenvoice modeling with sparse training data," *Speech and Audio Processing, IEEE Transactions on*, vol. 13, no. 3, pp. 345–354, 2005.

[3] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 4, pp. 788–798, 2011.

[4] N. Dehak, P. A. Torres-Carrasquillo, D. A. Reynolds, and R. Dehak, "Language recognition via i-vectors and dimensionality reduction." in *INTERSPEECH*, 2011, pp. 857–860.

[5] A. O. Hatch, S. S. Kajarekar, and A. Stolcke, "Within-class covariance normalization for svm-based speaker recognition." in *INTERSPEECH*, 2006.

[6] W. M. Campbell, D. E. Sturim, and D. A. Reynolds, "Support vector machines using gmm supervectors for speaker verification," *Signal Processing Letters, IEEE*, vol. 13, no. 5, pp. 308–311, 2006.

[7] P. Kenny, "Bayesian speaker verification with heavy tailed priors," in *Speaker and Language Recognition Workshop (IEEE Odyssey)*, 2010.

[8] P. Matejka, O. Glembek, F. Castaldo, M. J. Alam, O. Plchot, P. Kenny, L. Burget, and J. Cernocky, "Full-covariance ubm and heavy-tailed plda in i-vector speaker verification," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011, pp. 4828–4831.

[9] D. Martınez, O. Plchot, L. Burget, O. Glembek, and P. Matejka, "Language recognition in ivectors space," *Proceedings of Interspeech, Firenze, Italy*, pp. 861–864, 2011.

[10] P. Matějka, O. Plchot, M. Soufifar, O. Glembek, L. F. D'haro Enríquez, K. Veselỳ, F. Grézl, J. Ma, S. Matsoukas, and N. Dehak, "Patrol team language identification system for darpa rats p1 evaluation," 2012.

[11] K. J. Han, S. Ganapathy, M. Li, M. K. Omar, and S. Narayanan, "Trap language identification system for rats phase ii evaluation," in *Proc. Interspeech*, 2013.

[12] A. Kanagasundaram, R. Vogt, D. B. Dean, S. Sridharan, and M. W. Mason, "I-vector based speaker recognition on short utterances," in *Proceedings of the 12th Annual Conference of the International Speech Communication Association*. International Speech Communication Association (ISCA), 2011, pp. 2341–2344.

[13] A. Larcher, P. Bousquet, K. A. Lee, D. Matrouf, H. Li, and J.-F. Bonastre, "I-vectors in the context of phonetically-constrained short utterances for speaker verification," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 4773–4776.

[14] A. Lawson, M. McLaren, Y. Lei, V. Mitra, N. Scheffer, L. Ferrer, and M. Graciarena, "Improving language identification robustness to highly channel-degraded speech through multiple system fusion," 2013.

[15] M. Van Segbroeck, R. Travadi, and S. S. Narayanan, "Ubm fused total variability modeling for language identification," in *INTERSPEECH*, 2014, submitted.

[16] M. Li and S. Narayanan, "Simplified supervised i-vector modeling with application to robust and efficient language identification and speaker verification," *Computer, Speech, and Language*, 2014.

[17] O. Glembek, L. Burget, P. Matejka, M. Karafiát, and P. Kenny, "Simplification and optimization of i-vector extraction," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011, pp. 4516–4519.

[18] K. Walker and S. Strassel, "The rats radio traffic collection system," in *Proc. Odyssey*, 2012.