NORMALIZATION OF TOTAL VARIABILITY MATRIX FOR I-VECTOR/PLDA SPEAKER VERIFICATION

Wei Rao[1], *Man-Wai Mak*[1], *and Kong-Aik Lee*[2]

[1]Dept. of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong
[2]Human Language Technology Department, Institute for Infocomm Research, A*STAR, Singapore
ellen.wei-rao@connect.polyu.hk, enmwmak@polyu.edu.hk, kalee@i2r.a-star.edu.sg

ABSTRACT

Gaussian PLDA with uncertainty propagation is effective for i-vector based speaker verification. The idea is to propagate the uncertainty of i-vectors caused by the duration variability of utterances to the PLDA model. However, a limitation of the method is the difficulty of performing length normalization on the posterior covariance matrix of an i-vector. This paper proposes a method to avoid performing length normalization on i-vectors in Gaussian PLDA modeling so that uncertainty propagation can be directly applied without transforming the posterior covariance matrices of i-vectors. Instead of performing length normalization on i-vectors independently, the proposed method normalizes the column vectors of the total variability matrix. Because the i-vectors of all utterances are derived from the same normalized total variability matrix, they will be subject to the same degree of normalization, thereby avoiding the undesirable distortion introduced by the utterance-dependent length-normalization process. Experimental results on both NIST 2010 and 2012 SREs demonstrate that the proposed method achieves a performance similar to (and in some situations better than) that of Gaussian PLDA with length normalization. The method has the potential of improving the performance of uncertainty propagation for i-vector/PLDA speaker verification.

Index Terms— Total variability matrix, i-vectors, probabilistic linear discriminant analysis, uncertainty propagation, speaker verification.

1. INTRODUCTION

The I-vector/PLDA framework [1–4] is a state-of-the-art approach to speaker verification. However, this framework ignores the fact that i-vectors estimated from short utterances are less reliable than the ones estimated from long utterances. As a result, all i-vectors are considered equally reliable. When the utterance is short, the posterior covariance of the corresponding i-vector will become large, causing greater uncertainty in the estimated i-vector [5]. Kenny et al. [5] proposed a method called uncertainty propagation that propagates the uncertainty associated with the point estimate of an i-vector to the PLDA generative model through the Cholesky decomposition of the posterior covariance of the i-vector. This method can be performed on both Gaussian PLDA and heavy-tailed PLDA, and studies have shown that uncertainty propagation can substantially improve the performance of PLDA models [5, 6].

It was found in [4] that to use Gaussian PLDA for suppressing session and channel variability, the i-vectors should be subject to length normalization. However, length normalization is a nonlinear transformation, which causes difficulty for uncertainly propagation because it is not sure how to perform length normalization on the posterior covariance matrix of an i-vector [5]. To overcome this difficult, Kenny et al. [5] proposed two "unscented" transformation methods as a means to perform length normalization on the posterior covariance and Stafylakis et al. [6] avoid the normalization by forcing the posterior covariance to be diagonal. However, the methods are rather heuristic and the assumption of diagonal covariance may be invalid.

This paper proposes a method to avoid performing length normalization on i-vectors in Gaussian PLDA modeling without degrading performance. In other words, the method can achieve a similar performance as if length normalization has been performed. Because length normalization becomes unnecessary, how to perform length normalization on the posterior covariance in uncertainty propagation is not an issue anymore. Therefore, our method has the potential of improving the performance of systems that make use of uncertainty propagation.

Instead of performing length normalization on i-vectors independently, our method normalizes the column vectors of the total variability matrix $\mathbf{T}$. This has the effect of reducing the range of the variance in different dimensions of the subspaces defined by $\mathbf{T}$. The normalization can be done within each iteration of the EM algorithm for estimating $\mathbf{T}$. Alternatively, it can be done after the EM algorithm has completed. Unlike the conventional length normalization, this method will not introduce utterance-dependent non-linear distortion to the i-vectors. While there may be some distortion to the total variability matrix, all i-vectors will be subject to the same distortion because all of them are derived from the same T-matrix.

This paper is organized as follows. Section 2 reviews the i-vector extraction process and introduces the idea of normalizing the total variability matrix. Sections 3 and 4 outline the experiments that enable us to study the property of the normalized total variability matrix, and section 5 concludes our findings and highlights some potential future work.

2. NORMALIZATION OF TOTAL VARIABILITY MATRIX

2.1. I-vector Extraction

The i-vector approach is based on the idea of joint factor analysis (JFA) [1]. In [2], Dehak et al. notice that the channel factors in JFA also contain speaker-dependent information. This finding motivates them to model the total variability space (including channels and speakers) instead of modeling the channel- and speaker-spaces separately. Given an utterance of speaker s, the speaker- and channel-

This work was in part supported by The Hong Kong Research Grant Council, PolyU Grant No. G-YN18 and G-YL78.

dependent GMM-supervector [7] $\boldsymbol{m}_s$ is written as:

$$\boldsymbol{m}_s = \boldsymbol{m} + \mathbf{T}\mathbf{w}_s \tag{1}$$

where $\boldsymbol{m}$ is the GMM-supervector of the universal background model (UBM) [8] which is speaker- and channel-independent, $\mathbf{T}$ is a low-rank total variability matrix, and the posterior mean of $\mathbf{w}_s$ is a low-dimension vector called i-vector.

Given an utterance with D-dimensional acoustic vector sequence $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_T\}$ belonging to speaker s and an UBM $\mathbf{\Lambda}^{(b)} = \{\lambda_i^{(b)}, \boldsymbol{\mu}_i^{(b)}, \mathbf{\Sigma}_i^{(b)}\}_{i=1}^M$ with M mixture components, the zero-order and centered first-order Baum-Welch statistics are computed as follows [2,9,10]:

$$\mathbf{N}_{s,i} = \sum_{t=1}^{T} \Pr(i|\mathbf{x}_t) \text{ and } \tilde{\mathbf{F}}_{s,i} = \sum_{t=1}^{T} \Pr(i|\mathbf{x}_t)(\mathbf{x}_t - \boldsymbol{\mu}_i^{(b)}) \tag{2}$$

where

$$\Pr(i|\mathbf{x}_t) = \frac{\lambda_i^{(b)} \mathcal{N}(\mathbf{x}_t; \boldsymbol{\mu}_i^{(b)}, \mathbf{\Sigma}_i^{(b)})}{\sum_{j=1}^{M} \lambda_j^{(b)} \mathcal{N}(\mathbf{x}_t; \boldsymbol{\mu}_j^{(b)}, \mathbf{\Sigma}_j^{(b)})}, \quad i = 1, \ldots, M.$$

is the posterior probability of mixture components i given $\mathbf{x}_t$. The posterior covariance and posterior mean associated with an i-vector are given by:

$$\mathrm{Cov}(\mathbf{w}_s, \mathbf{w}_s) = \mathbf{L}_s^{-1} \tag{3}$$

$$\mathbf{w}_s = \mathbf{L}_s^{-1} \mathbf{T}^{\mathsf{T}} {\mathbf{\Sigma}^{(b)}}^{-1} \tilde{\mathbf{F}}_s \tag{4}$$

where

$$\mathbf{L}_s = \mathbf{I} + \mathbf{T}^{\mathsf{T}} {\mathbf{\Sigma}^{(b)}}^{-1} \mathbf{N}_s \mathbf{T} \tag{5}$$

is a precision matrix and $\mathbf{I}$ is the identity matrix. $\mathbf{N}_s$ is an $MD \times MD$ diagonal matrix whose diagonal blocks are $\mathbf{N}_{s,i}\mathbf{I}$. $\tilde{\mathbf{F}}_s$ is an $MD \times 1$ supervector formed by concatenating the centered first-order Baum-Welch statistics $\tilde{\mathbf{F}}_{s,i}$. $\mathbf{\Sigma}^{(b)}$ is a covariance matrix modeling the residual variability not captured by the $MD \times R$ total variability matrix $\mathbf{T}$. In practice, we substitute this matrix by the covariance matrices of the UBM, i.e., $\mathbf{\Sigma}^{(b)} = \mathrm{diag}\{\mathbf{\Sigma}_1^{(b)}, \ldots, \mathbf{\Sigma}_M^{(b)}\}$. The posterior mean (Eq. 4) is the i-vector representing the speaker s.

The estimation of total variability matrix is almost identical to that of the eigenvoice matrix in JFA [9,11]. It is trained by maximizing a likelihood objective function via the EM algorithm. Because $\mathbf{N}_s$ is a diagonal matrix, The maximum likelihood update formula for $\mathbf{T}$ can be obtained from:

$$\mathbf{C}_i = \sum_s \tilde{\mathbf{F}}_{s,i} {\mathbf{w}_s}^{\mathsf{T}}, \tag{6}$$

$$\mathbf{A}_i = \sum_s \mathbf{N}_{s,i} \left(\mathbf{L}_s^{-1} + \mathbf{w}_s {\mathbf{w}_s}^{\mathsf{T}} \right), \tag{7}$$

$$\mathbf{T}_i = \mathbf{C}_i \mathbf{A}_i^{-1} \quad i = 1, \ldots, M \tag{8}$$

where $\mathbf{T}_i$ is a $D \times R$ submatrix of $\mathbf{T}$. In summary, the total variability matrix can be obtained by iteratively performing the E-step (Eq. 4 and Eq. 5) and M-step (Eq. 6 – Eq. 8).

2.2. Normalization of Total Variability Matrix

A problem in factor analysis [12] is unidentifiability of the factor loading matrix up to rotation and scaling [13–15], i.e., the factor loading matrix and latent factors cannot be uniquely identified. Because the theory behind i-vectors is based on factor analysis, the same problem will also occur in i-vector extraction. Let us use scaling as an example. Denote an $R \times R$ arbitrary diagonal matrix $\mathbf{\Phi}$ with non-zero diagonal elements. Then we define:

$$\begin{aligned} \tilde{\mathbf{T}} &= \mathbf{T}\mathbf{\Phi}^{-1} \\ \tilde{\mathbf{w}}_s &= \mathbf{\Phi}\mathbf{w}_s. \end{aligned} \tag{9}$$

Substituting Eq. 9 into Eq. 1 and noting that $\mathbf{\Phi}^{-1}\mathbf{\Phi} = \mathbf{I}$, we obtain

$$\begin{aligned} \tilde{\boldsymbol{m}}_s &= \boldsymbol{m} + \tilde{\mathbf{T}}\tilde{\mathbf{w}}_s \\ &= \boldsymbol{m} + \mathbf{T}\mathbf{\Phi}^{-1}\mathbf{\Phi}\mathbf{w}_s \\ &= \boldsymbol{m} + \mathbf{T}\mathbf{w}_s \\ &= \boldsymbol{m}_s \end{aligned} \tag{10}$$

Eq. 10 shows that $\tilde{\boldsymbol{m}}_s$ will be equal to $\boldsymbol{m}_s$ for an arbitrary scaling matrix $\mathbf{\Phi}$ as long as the matrix is diagonal and all of its elements are non-zero. Hence, the solutions of $\mathbf{T}$ and $\mathbf{w}_s$ are non-unique.

To use Gaussian PLDA, length normalization should be performed on i-vectors [4]. However length normalization is a non-linear transformation, which causes difficulty for uncertainty propagation because there is no straight forward way to perform length normalization on the posterior covariance of i-vectors. Inspired by the unidentifiability problem of factor analysis [13–15], we propose to perform column normalization of the total variability matrix instead of performing length normalization on i-vectors.

Denote $\mathbf{T} = [\mathbf{t}_1, \ldots, \mathbf{t}_R]$ where $\mathbf{t}_r$ is an $MD \times 1$ column vector. The normalized total variability matrix is defined as:

$$\mathbf{T}_{norm} = \left[\frac{\mathbf{t}_1}{\|\mathbf{t}_1\|} \cdots \frac{\mathbf{t}_R}{\|\mathbf{t}_R\|} \right] = \mathbf{T}\mathbf{\Phi}^{-1} \tag{11}$$

where $\mathbf{\Phi}$ is an $R \times R$ diagonal matrix and its diagonal element ϕ_r is the norm of the r-th column of $\mathbf{T}$. This formula can be applied after the whole EM algorithm has completed. Alternatively, it can be applied after the completion of each M-step (Eq. 6 – Eq. 8). In the sequel, we refer to this normalization process as **T-matrix normalization**.

Once $\mathbf{T}_{norm}$ has been obtained, it can be used for estimating the i-vectors of any utterances. Specifically, given an utterance, its i-vector can be obtained by Eq. 4 with $\mathbf{T}$ replaced by $\mathbf{T}_{norm}$. The property of T-matrix normalization will be discussed in Section 4.

3. EXPERIMENTAL SETUP

3.1. Speech Data and Acoustic Features

Experiments were performed on the male *extended core set* of NIST 2010 and male *core set* of NIST 2012 Speaker Recognition Evaluations (SREs). For NIST 2010 SRE [16], interview and microphone utterances of the extended core task were used, i.e., Common Conditions 1, 2, and 4. In the sequel, we use "CC" to denote common evaluation conditions. Male utterances from NIST 2005–2008 SREs were used as development data (UBM, total variability matrix, and PLDA). For NIST 2012 SRE [17], male phonecall utterances of the core task were used, i.e., CC2. The speech files of male speakers in NIST 2005–2010 SREs were used as development data for training the UBM, total variability matrix, and PLDA models.

A Voice activity detector [18, 19] was used to detect the speech regions of each utterance. 19 MFCCs together with energy plus their 1st- and 2nd-derivatives were extracted from the speech regions, followed by cepstral mean normalization [20] and feature warping [21] with a window size of 3 seconds. A 60-dim acoustic vector was extracted every 10ms, using a Hamming window of 25ms.

3.2. Total Variability Modeling and PLDA

The i-vector systems are based on a gender-dependent UBM with 1024 mixtures. For NIST 2010 SRE, 4,073 microphone utterances from NIST 2005–2008 SREs were used for training the microphone UBM. 9,561 microphone utterances from NIST 2005–2008 SREs were selected for estimating a microphone total variability matrix with 400 total factors. The same data set was used for estimating the Gaussian PLDA models.

For NIST 2012 SRE, 3,500 microphone utterances and 3,501 telephone utterances from NIST 2005–2008 SREs were used for training a channel-independent UBM. We selected 14,875 telephone and interview conversations from 575 speakers in NIST 2006–2010 SREs to estimate a total variability matrix with 400 total factors. 15,662 telephone and interview conversations from 673 speakers in NIST 2006–2010 SREs were used for training Gaussian PLDA models with 150 latent variables.

We compared the performance of i-vectors extracted from the ordinary total variability matrices with that from the normalized total variability matrices. For each case, we also investigate the performance of systems with and without whitening (using within-class covariance normalization [22]) and i-vector length normalization [4].

4. RESULTS AND DISCUSSIONS

4.1. Property of Normalized Total Variability Matrix

If T-matrix normalization (Eq. 11) is applied after the completion of each M-step (Eq. 6 – Eq. 8), it may affect the estimated i-vectors in the E-step, which may in turn affect the convergency of the EM algorithm. It is of interest to investigate the effect of T-matrix normalization on the convergency of the EM algorithm. To this end, 5,000 utterances from NIST 2006–2010 SREs were selected for training two total variability matrices with 100 total factors: one without T-matrix normalization and another one with T-matrix normalization. The training of both matrices was started with the same initial random matrix. Fig. 1 shows the log-likelihood (Proposition 2 of [11]) against the EM iteration numbers. Evidently, performing T-matrix normalization within the EM iterations will not affect the convergency of the EM algorithm.

As introduced in Section 2.1, the posterior covariance of i-vector is the inverse of the precision matrix $\mathbf{L}_s$. It is of interest to explore the effect of T-matrix normalization on the numerical stability of the posterior covariance. To this end, the same experimental setup as in Fig. 1 was adopted. Fig. 2 shows the condition number[1] of the precision matrix $\mathbf{L}_s$ against the EM iteration numbers. Evidently, with T-matrix normalization, the condition numbers are much smaller than those without T-matrix normalization. In addition, without T-matrix normalization, the condition number of $\mathbf{L}_s$ starts to increase gradually after 6 iterations. On the other hand, with T-matrix normalization, the condition number keeps on decreasing, suggesting that the T-matrix normalization helps improve the numerical stability of the posterior covariance matrices.

4.2. Performance Comparison

Table 1 and Table 2 show the performance of T-matrix normalization in NIST 2010 and 2012 SREs, respectively. The results show that without T-matrix normalization (Sys. 1–4), i-vector length normalization is very critical to the performance of i-vector/PLDA systems. However, with T-matrix normalization (Sys. 5, 6, 9, and 10), i-vector length normalization is not necessary and the performance is almost identical to and sometimes better than the best performing baseline (Sys. 4). As for the whitening of i-vectors (WCCN), its effect on the i-vectors derived from the normalized T-matrix is inconclusive. As shown in Sys. 6 and Sys. 10 of Table 1 and Table 2, WCCN reduces the EER but it also increases the minimum DCF for some conditions. In addition, results of Sys. 7, 8, 11, and 12 in Table 1 and Table 2 further show that i-vector length normalization is not necessary whenever the total variability matrix has been normalized. Comparisons among Sys. 5, 6, 9, and 10 in Table 1 and Table 2 reveal that it is better to apply T-matrix normalization after the EM algorithm has completed. This is in fact an advantage of the pro-

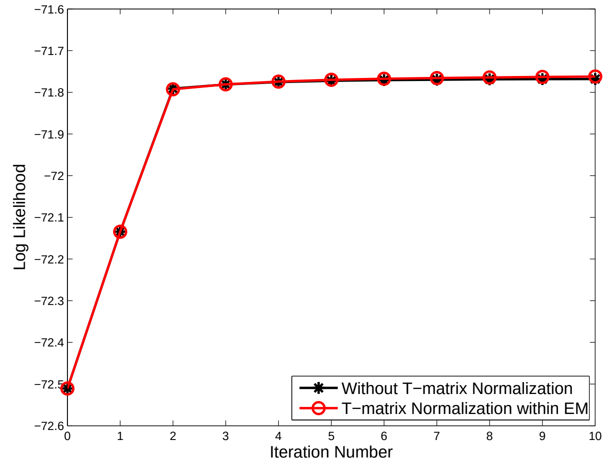

Fig. 1. Log-likelihood against the EM iteration number. See the caption of Table 1 for the interpretations of the legend.

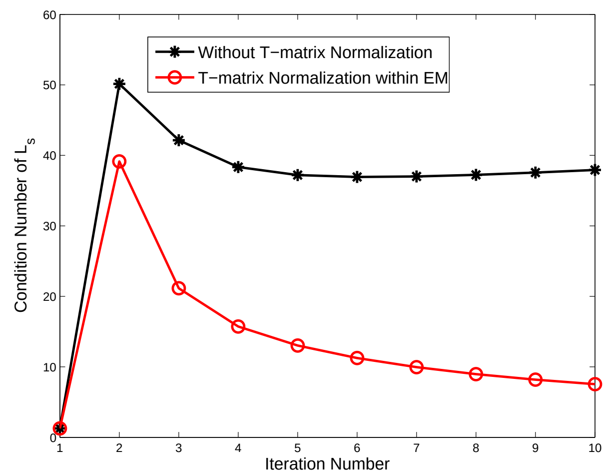

Fig. 2. Condition number of precision matrix $\mathbf{L}_s$ in Eq. 5 against the EM iteration number. See the caption of Table 1 for the interpretations of the legend.

[1]The condition number of a matrix can be used to measure the numerical stability of matrix inversion. The smaller the condition number (the closer to 1), the better is the condition of the matrix [23].

Sys.	T-matrix Normalization	I-vector Preprocessing	EER (%)			MinNDCF(2010)		
			CC1	CC2	CC4	CC1	CC2	CC4
1	None	None	4.04	5.64	5.30	0.542	0.727	0.590
2	None	WCCN	1.97	3.23	3.66	0.364	0.534	0.519
3	None	LN	2.22	3.56	3.69	0.373	0.532	0.568
4	None	WCCN+LN	1.52	2.41	2.97	0.308	0.457	0.442
5	Within EM	None	1.59	2.61	3.10	0.255	0.433	0.445
6	Within EM	WCCN	1.41	2.43	3.27	0.354	0.515	0.522
7	Within EM	LN	2.83	4.01	3.92	0.402	0.562	0.574
8	Within EM	WCCN+LN	1.42	2.45	**2.90**	0.290	0.445	0.436
9	After EM	None	1.50	2.49	3.07	**0.233**	**0.414**	**0.433**
10	After EM	WCCN	**1.37**	**2.34**	3.23	0.353	0.521	0.520
11	After EM	LN	3.13	4.33	4.24	0.427	0.580	0.582
12	After EM	WCCN+LN	1.57	2.56	2.97	0.292	0.451	0.436

Table 1. The performance of i-vector/PLDA based speaker verification with and without normalization of total variability matrix for NIST 2010 SRE (male speakers) under the common conditions involving microphone recordings. "Sys.": systems. "T-matrix": total variability matrix. "EM": expectation maximization iteration. "T-matrix Normalization within EM": applying the normalization of total variability matrix after the completion of each M-step (Eq. 6 – Eq. 8). "T-matrix Normalization after EM": applying the normalization of total variability matrix after the whole EM algorithm has completed. The methods in i-vector preprocessing are named by the processes applied to the i-vectors for computing the verification scores. For example, *WCCN+LN* means that whitening and length normalization were performed on i-vectors before training the PLDA model.

posed method, because only a single normalization step is needed.

Fig. 3 shows the DET performance of systems with and without T-matrix normalization. It also suggests that i-vector length normalization is not necessary whenever the total variability matrix has been normalized.

Sys.	T-matrix Normalization	I-vector Preprocessing	EER (%)	MinNDCF (2012)
1	None	None	11.74	0.567
2	None	WCCN	4.17	0.472
3	None	LN	6.04	0.351
4	None	WCCN+LN	2.82	0.351
5	Within EM	None	3.67	0.286
6	Within EM	WCCN	3.35	0.287
7	Within EM	LN	4.75	0.464
8	Within EM	WCCN+LN	**2.72**	0.351
9	After EM	None	3.30	0.273
10	After EM	WCCN	3.25	**0.272**
11	After EM	LN	5.86	0.538
12	After EM	WCCN+LN	3.36	0.394

Table 2. The performance of i-vector/PLDA based speaker verification with and without normalization of total variability matrix for common condition 2 of NIST 2012 SRE (male speakers). The interpretations of methods in the 2nd and 3rd columns are described in the caption of Table 1.

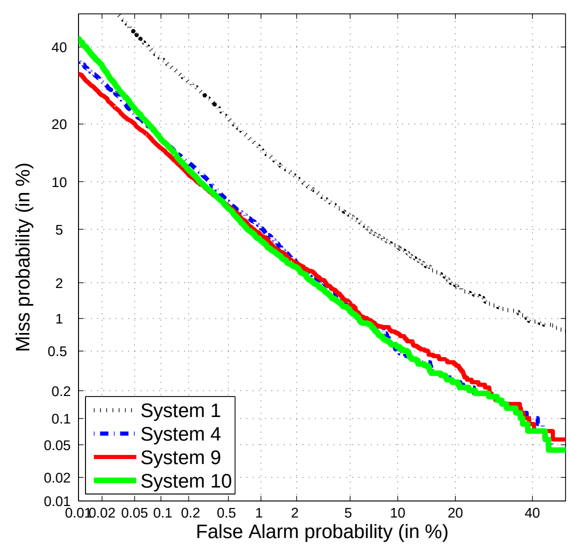

Fig. 3. The DET performance under CC2 in NIST 2010 SRE. See Table 1 for the nomenclature of methods in the legend.

5. CONCLUSIONS AND FUTURE WORK

Inspired by the unidentifiability problem of factor analysis, this paper proposes performing normalization on the total variability matrix instead of performing length normalization on i-vectors in i-vector/PLDA speaker verification. Experimental results show that the proposed T-matrix normalization can achieve almost the same performance as i-vector length normalization. The significance of the results is that the method opens up opportunity to enhance the performance of uncertainly propagation because the issue of applying length normalization on the posterior covariance of i-vectors can now be resolved. The T-matrix normalization process imposes a tight constraint on the total variability matrix (norms of all columns equal to 1), whereas the conventional T-matrix estimation imposes no constraint. For future work, it is of interest to investigate the effect of relaxing such constraint on system performance.

6. REFERENCES

[1] P. Kenny, G. Boulianne, P. Ouellet, and P. Dumouchel, "Joint factor analysis versus eigenchannels in speaker recognition," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 15, no. 4, pp. 1435–1447, May 2007.

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

[3] P. Kenny, "Bayesian speaker verification with heavy-tailed priors," in *Proc. of Odyssey: Speaker and Language Recognition Workshop*, Brno, Czech Republic, Jun. 2010.

[4] D. Garcia-Romero and C. Y. Espy-Wilson, "Analysis of i-vector length normalization in speaker recognition systems," in *Proc. of Interspeech 2011*, Florence, Italy, Aug. 2011, pp. 249–252.

[5] P. Kenny, T. Stafylakis, P. Ouellet, M. J. Alam, and P. Dumouchel, "PLDA for speaker verification with utterances of arbitrary duration," in *Proc. ICASSP 2013*, Vancouver, Canada, May 2013, pp. 7649–7653.

[6] T. Stafylakis, P. Kenny, P. Ouellet, J. Perez, M. Kockmann, and P. Dumouchel, "Text-dependent speaker recogntion using PLDA with uncertainty propagation," in *Proc. Interspeech 2013*, Lyon, France, Aug. 2013.

[7] 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," in *Proc. ICASSP*, Toulouse, France, May 2006, vol. 1, pp. 97–100.

[8] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing*, vol. 10, no. 1–3, pp. 19–41, Jan. 2000.

[9] P. Kenny, P. Ouellet, N. Dehak, V. Gupta, and P. Dumouchel, "A study of inter-speaker variability in speaker verification," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 16, no. 5, pp. 980–988, 2008.

[10] P. Kenny, "A small foot-print i-vector extractor," in *Proc. Odyssey 2012*, 2012.

[11] P. Kenny, G. Boulianne, and P. Dumouchel, "Eigenvoice modeling with sparse training data," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 3, pp. 345–359, May 2005.

[12] S. J. D. Prince, *Computer Vision: Models Learning and Inference*, Cambridge University Press, 2012.

[13] K. P. Murphy, *Machine learning: a probabilistic perspective*, MIT press, 2012.

[14] "www.genome.duke.edu/labs/engelhardt/courses/sta561.html," Lecture Note from Duke University, 2013.

[15] H. F. Kaiser, "The varimax criterion for analytic rotation in factor analysis," *Psychometrika*, vol. 23, no. 3, pp. 187–200, 1958.

[16] 2010 NIST Speaker Recognition Evaluation, "http://www.itl.nist.gov/iad/mig//tests/sre/2010/," 2010.

[17] 2012 NIST Speaker Recognition Evaluation, "http://www.nist.gov/itl/iad/mig/sre12.cfm," 2012.

[18] H.B. Yu and M.W. Mak, "Comparison of voice activity detectors for interview speech in NIST speaker recognition evaluation," in *Proc. of Interspeech 2011*, Florence, Aug. 2011, pp. 2353–2356.

[19] M.W. Mak and H.B. Yu, "A study of voice activity detection techniques for NIST speaker recognition evaluations," *Computer Speech and Language*, vol. 28, no. 1, pp. 295 – 313, Jan. 2014.

[20] B. S. Atal, "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification," *J. Acoust. Soc. Am.*, vol. 55, no. 6, pp. 1304–1312, Jun. 1974.

[21] J. Pelecanos and S. Sridharan, "Feature warping for robust speaker verification," in *Proc. of Odyssey: Speaker and Language Recognition Workshop*, Crete, Greece, Jun. 2001, pp. 213–218.

[22] A. Hatch, S. Kajarekar, and A. Stolcke, "Within-class covariance normalization for SVM-based speaker recognition," in *Proc. of the 9th International Conference on Spoken Language Processing*, Pittsburgh, PA, USA, Sep. 2006, pp. 1471–1474.

[23] MathWorks, "www.mathworks.com/help/matlab/ref/cond.html," 2014.