INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Generalized Discriminant Analysis (GDA) for Improved i-Vector Based Speaker Recognition

Fahimeh Bahmaninezhad, John H.L. Hansen

Center for Robust Speech Systems (CRSS), University of Texas at Dallas, Richardson, TX 75080
{fahimeh.bahmaninezhad,john.hansen}@utdallas.edu

Abstract

In general, the majority of recent speaker recognition systems employ an i-Vector configuration as their front-end. Post-processing of i-Vectors usually requires a Linear Discriminant Analysis (LDA) phase to reduce the dimensions of the i-Vectors as well as improve discrimination of speaker classes based on the Fisher criterion. Given that channel, noise, and other types of mismatch are generally present in the data, it is better to discriminate the speaker's data non-linearly. Generalized Discriminant Analysis (GDA) uses kernel functions to map the data into a high dimensional feature-space which leads to non-linear discriminant analysis. In this study, we replace LDA with GDA in an i-Vector based speaker recognition system and study the effectiveness of various kernel functions. It is shown, based on equal error rate (EER) and minimum of detection cost function, that GDA not only improves performance for regular test utterances, but is also useful for short duration test segments. NIST2010 Speaker Recognition Evaluation (SRE) core and extended-core (coreext) conditions are employed for experiments; in addition, we evaluate the system for short duration segments on the 10-sec test condition and truncated coreext test data. The relative improvement in EER is 20% for the cosine kernel employed here with GDA processing.

Index Terms: Speaker recognition, i-Vector, Generalized discriminant analysis, Kernel.

1. Introduction

Speaker recognition systems using i-Vector feature representation [1] and Probabilistic Linear Discriminant Analysis (PLDA) [2] scoring has been widely used as the state-of-the-art [3, 4]. These systems performe well in clean and channel mismatch conditions; especially for core conditions of NIST speaker recognition evaluation sets [1, 5]. Data introduced in these challenges, and also present in real world applications, may contain distortion such as noise, speaker or channel variations. In general, LDA [6] has been applied to i-Vectors as a post-processing phase to minimize the ratio of within class to between class covariance, and separate speaker-dependent factors. However, discriminating speaker classes may be optimized by employing non-linear discriminant analysis methods, when data are not clean.

LDA is a typical post-processing statistical method for classification and pattern recognition problems. It finds the exact optimum solution and performs well for linear problems [6]. LDA assumes classes are normally distributed, share the same covariance, and can be discriminated linearly [7]. To, remove these restricting assumptions, researchers have proposed variations of LDA, such as: heteroscedastic discriminant analysis [8], generalized discriminant analysis or kernel DA [7, 9], mixture discriminant analysis [10], etc. In this study, we consider GDA in order to remove the assumption of linear separability for classes.

In [11, 12], the application of various kernels in Support Vector Machines (SVMs) with Joint Factor Analysis (JFA) [13] has been studied, which show that the cascade of Within Class Covariance Normalization (WCCN) [14] (as a method of compensating for residual channel aspects present in the speaker factor space) and cosine kernel results in the best combined performance. When i-Vector feature representation introduced, the kernel SVMs and cosine scoring were the main classification approaches [1]. After that, the combination of i-Vector and PLDA verification became the most popular speaker recognition system. To improve i-Vector/PLDA, [15] applied non-parametric discriminant analysis or nearest-neighbor discriminant analysis (NDA) approach instead of traditional LDA and proved NDA is effective especially in channel degraded and noisy conditions. Also, [16] propagates the uncertainty of i-Vectors in LDA to overcome the problem of short duration test utterances in speaker recognition systems. Here, the effectiveness of kernel based discriminant analysis approaches in the i-vector/PLDA system is studied.

Many studies attempt to make i-Vector based systems robust to noise [17], or short duration utterances [18, 19]. In these cases, extracted i-Vectors are not as reliable as before; therefore, the uncertainty of i-Vectors has been propagated through the system, or different score calibration methods [18] have been introduced to partially address the problem. Here, we aim to study the effectiveness of GDA for long and short duration test segments on NIST SRE 2010 [20] task that already spans a range of distortions. We will assess i-Vector/PLDA system when GDA post-processed i-Vectors are employed.

In this paper, first Sec. 2 presents an overview of i-Vector based speaker recognition systems. Next, Sec. 3 covers conventional LDA and GDA approaches. In Sec. 4, the experimental setup and results comparing LDA and GDA will be provided. Conclusions and future work are summarized in Sec. 5.

2. i-Vector based speaker recognition

The overall block-diagram of our i-Vector based speaker recognition system using PLDA scoring is depicted in Figure 1. In an i-Vector configuration, the speaker and channel dependent

This project was funded by AFRL under contract FA8750-15-1-0205 and partially by the University of Texas at Dallas from the Distinguished University Chair in Telecommunications Engineering held by J. H. L. Hansen.

http://dx.doi.org/10.21437/Interspeech.2016-1523

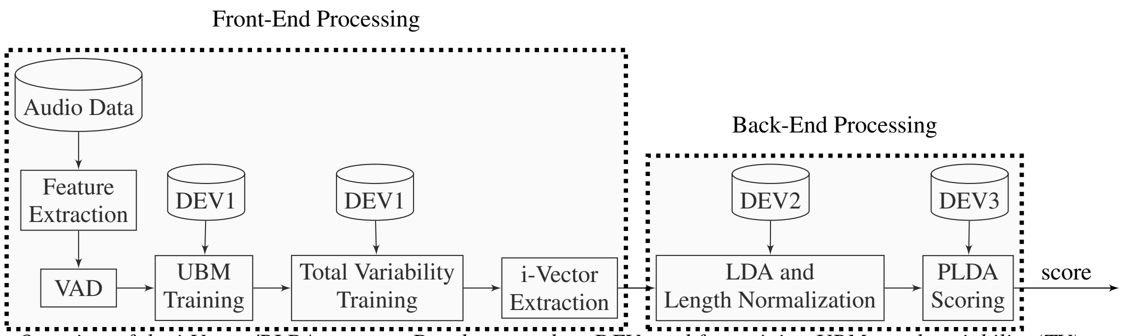

Figure 1: *Overview of the i-Vector/PLDA systems. Development data DEV used for training UBM, total variability (TV) matrix, LDA (or GDA) and PLDA, enrollment and test data used for evaluation of the system.*

Gaussian Mixture Model (GMM) supervector is factorized as:

$$M = m + Tw \tag{1}$$

where m is the mean supervector of Universal Background Model (UBM), T is the low rank total variability matrix, and w is identity vector well known as i-Vector.

Both UBM and total variability matrix are trained with the Expectation Maximization (EM) algorithm using development data. In the E-step, w is considered as a latent variable and is assumed to have a standard normal prior distribution [1, 21]. After training the UBM and total variability matrix, i-Vectors are extracted as the mean of posterior distribution of w that is equal to [1, 21]:

$$\hat{w}(u) = (I + T^T\Sigma^{-1}N(u)T)^{-1}T^T\Sigma^{-1}S(u) \tag{2}$$

where Σ is the covariance matrix of UBM, $N(u)$ and $S(u)$ are zeroth and centralized first order statistics of the UBM extracted from utterance u. Finally, the post-processed i-Vectors are used to calculate the PLDA score.

3. Post-processing of i-Vectors

Post-processing of i-Vectors usually contains length normalization, applying LDA and WCCN [16]. In this study, we explore the performance of speaker recognition when LDA is replaced with GDA. The following subsections briefly discuss traditional LDA versus GDA method.

3.1. Linear discriminant analysis (LDA)

LDA finds a linear transformation of features that maximizes the Fisher-Rao criterion. The separation of speaker classes in the direction of W is equal to,

$$\lambda = \frac{W^T S_B W}{W^T S_W W}, \tag{3}$$

where S_B and S_W represents the between class and within class scatters respectively. When W maximizes $S_W^{-1}S_B$, the class separation will be maximized as well. In other words, the eigenvectors corresponding to the largest eigenvalues in solving $\lambda S_W W = S_B W$ leads to the optimal projection matrix W.

For dimensionality reduction to k, the eigenvectors of the k largest eigenvalues are placed in matrix W. Thereafter, the projected feature vectors are calculated by $W^T x$, where x represents the input feature vector.

In Eq. 3, S_B and S_W are defined as,

$$S_B = \frac{1}{C}\sum_{c=1}^{C} n_c(\mu_c - \mu)(\mu_c - \mu)^T \tag{4}$$

$$S_W = \frac{1}{C}\sum_{c=1}^{C}\sum_{k\in c}(x_k - \mu_c)(x_k - \mu_c)^T, \tag{5}$$

where C represents the number of speaker classes and n_c represents the number of samples in class c. In addition, μ_c and μ are the mean of class c and overall mean of samples, respectively.

3.2. Generalized discriminant analysis (GDA)

Traditional LDA assumes data are normally distributed and distinct classes share the same covariance matrix; then it finds a linear transformation of the feature vectors. On the other hand, GDA first maps the data into a new feature space and then finds a linear transformation. Mapping to the new space is carried out using kernel methods. As the mapped feature vectors are nonlinearly related to the input versions, GDA effectively provides a non-linear discriminant analysis for the input feature data [7].

More specifically, GDA first maps feature vectors x in space X to feature vectors $\phi(x)$ in space F. Next, the between and within class scatters will be updated as (assuming observations are centered in F),

$$S_{Bf} = \frac{1}{C}\sum_{c=1}^{C} n_c\bar{\phi}_c\bar{\phi}_c^{\,T} \tag{6}$$

$$S_{Wf} = \frac{1}{C}\sum_{c=1}^{C}\sum_{k\in c}\phi(x_k)\phi(x_k)^T \tag{7}$$

where $\bar{\phi}_c$ is the mean of class c in feature space F (i.e. mean of $\phi(x)$ for x in class c). To generalize LDA, we need to formulate the eigenvalue resolution problem in a dot-product format. Let us define the following kernel function:

$$k(x_i, x_j) = \phi(x_i)^T\phi(x_j) \tag{8}$$

where i and j range from 1 to the total number of training samples, (i.e., n_x). Then, define K to be a $n_x \times n_x$ matrix containing $k(x_i, x_j)$. By defining the block-diagonal matrix $M = (M_c)_{c=1,\ldots,C}$ with the same size as K for each $M_c = \frac{1}{n_c} \times I(n_c \times n_c)$; then Eq. 3 in the feature space F can be formulated as,

$$\lambda_f = \frac{\alpha^T KMK\alpha}{\alpha^T KK\alpha}, \tag{9}$$

where α are the coefficient vectors that satisfy $\nu = \sum_{c=1}^{C}\sum_{k\in c}\alpha_k\phi(x_k)$; ν are the eigenvectors of $\lambda_f S_{Wf}\nu =$

Table 1: *Number of speakers (Spkrs) and segments used for training the models of UBM, TV, LDA/GDA, PLDA, and the number of enrollment segments and statistics of trails.*

Enrollment/Test	UBM-TV		PLDA		LDA/GDA		Enrollment	Trials	
combination	Spkrs	Segments	Spkrs	Segments	Spkrs	Segments	Spkrs	Target	nonTarget
Core/Core							2426	353	13707
Core/10Sec	5756	57273	1115	13605	1078	10000	2426	290	11700
Coreext/3,5,10,20,40s							5237	3465	175873

$S_{Bf}\nu$. Since, the eigenvectors are linear combinations of feature vectors in space F, there exist a non-unique set of α coefficients. More details are provided in [7].

To solve Eq. 9, matrix K can be decomposed as:

$$K = P\Gamma P^T. \tag{10}$$

By defining $\beta = \Gamma P^T \alpha$ and replacing K with Eq. 10 in Eq. 9 and simplifying the equation, we can reach the following eigenvector system:

$$\lambda_f \beta = P^T M P \beta. \tag{11}$$

For eigenvectors β, there exists $\alpha = P\Gamma^{-1}\beta$. From α, the eigenvectors ν can be computed which leads to the projection matrix in feature space F.

4. Experiments and results

4.1. Database and experimental setup

For all systems in the experiments, 60 dimensional Mel-frequency features have been extracted, that include 19 dimensional static features as well as the frame energy along with their delta and delta-delta coefficients. Speech signals have been framed using 25-ms length windows with a 10-ms skip rate. In addition, features have been normalized using a 3-sec sliding window. Next, energy-based voice activity detection (VAD) was used to remove non-speech frames.

Experiments have been carried out on NIST SRE2010 [20], telephone condition (condition 5). 2048-mixture UBM and total variability matrix have been trained using data collected from SRE2004, 2005, 2006, 2008, and Switchboard II Phase 2 and 3, and Switchboard Cellular Part 1 and 2 (both male and female speakers). Next, 600-dimensional i-Vectors were extracted for all utterances. For LDA and GDA, the dimension size is reduced to 400, followed by length normalization. Training data for LDA, GDA, and PLDA is restricted to male speakers from NIST SRE2004, 2005, 2006, 2008 data. In addition, the trails used for experiments just contain male enrollment and test segments.

The enrollment/test segment condition combinations that have been evaluated in this study, the number of speakers and segments for training UBM, total variability matrix, LDA, GDA and PLDA, data used for enrollment, and statistics of trials are summarized in Table 1. Core and extended core conditions (coreext) have duration ranging between 3 to 5 minutes.

To examine the effectiveness of GDA for short test segments, the coreext test data was truncated into 3-sec, 5-sec, 10-sec, 20-sec, and 40-sec segments. Extracting these short test data has been carried out after applying VAD; therefore, they do not contain non-speech frames. In addition, no modifications have been applied on the enrollment or training data.

4.2. Kernel variations for GDA

The various kernel functions that have been used in the experiments are covered in this subsection. The linear kernel leads to traditional LDA, where between the i-Vectors w_1 and w_2 is defined as,

$$k(w_1, w_2) = \langle w_1, w_2 \rangle. \tag{12}$$

We use this as a baseline discriminant analysis. The cosine kernel is also used and defined as,

$$k(w_1, w_2) = \frac{\langle w_1, w_2 \rangle}{\| w_1 \| \| w_2 \|}. \tag{13}$$

The angles between i-Vectors is the only aspect captured by the cosine kernel. [1] states that the magnitude of i-Vectors may just contain information about channel and session which is not valuable in speaker recognition; therefore, when the cosine kernel removes the magnitude, we expect an improvement over a linear kernel.

The Within Class Covariance Normalization (WCCN) suppresses channel affects without removing any direction in the feature space. The projection matrix B for WCCN is achieved by a Cholesky decomposition of the within class scatter in Eq. 5 as $S_w^{-1} = BB^T$. Here, we apply WCCN to the cosine kernel that updates it as (this kernel will be referred to "WCCN-Cosine" kernel in the experiments):

$$k(w_1, w_2) = \frac{(B^T w_1)^T (B^T w_2)}{\sqrt{(B^T w_1)^T (B^T w_1)} \sqrt{(B^T w_2)^T (B^T w_2)}}. \tag{14}$$

The other kernel variation (named as "LDA-Cosine") uses the LDA projection matrix in the cosine kernel. The background on LDA was provided in Sec. 3.1. If we name the projection matrix as A, which is the ordered eigenvectors based on the highest values of eigenvalues, then the kernel would be,

$$k(w_1, w_2) = \frac{(A^T w_1)^T (A^T w_2)}{\sqrt{(A^T w_1)^T (A^T w_1)} \sqrt{(A^T w_2)^T (A^T w_2)}}. \tag{15}$$

Here, we have extracted 600-dimensional i-Vectors, and for this kernel the eigenvectors of the 600 largest eigenvalues have been selected for the projection matrix A. Therefore, we did not reduce the dimension of i-Vectors with projection matrix A; however, after transforming with A and applying cosine kernel the dimension is reduced to 400, as the other kernels.

We also used the cascade of LDA and WCCN to project the feature vectors, and then employ the cosine kernel. LDA and WCCN have different objectives in finding the projection matrix; therefore, we examined their combination in the application of kernel for GDA. We will refer to the kernel as "LDA-WCCN-Cosine" kernel.

In [22], the authors proposed Gaussianized Cosine Distance Scoring (GCDS) that improved traditional cosine distance scoring. It was claimed there that estimating WCCN projection matrix in noisy and/or channel mismatch condition is difficult. Therefore, they replaced the cascade of LDA, WCCN, and cosine distance scoring with GCDS method. Here, we take advantage of this idea and modify the algorithm to be used as a kernel function. Therefore, our "Gaussianized cosine kernel" is based on the the following routine:

Table 2: *Speaker recognition results comparing LDA and GDA using various kernel functions in terms of EER/minDCF. LDA and GDA reduce the dimension of i-Vectors from 600 to 400.*

	LDA	**GDA**				
Enrollment/Test Segments	Linear	Cosine	WCCN-Cosine	LDA-Cosine	LDA-WCCN-Cosine	Gaussianized-Cosine
Core/Core	1.416/.0353	**1.133**/.0351	1.539/.0346	1.416/.0337	1.641/**.0289**	**1.133**/.0339
Core/10Sec	4.838/.0586	5.172/.0624	5.517/**.0575**	**4.828**/.0621	4.854/.0671	**4.828**/.0596
Coreext/Coreext	1.438/**.0301**	**1.384**/.0309	1.558/.0319	1.44/.0320	1.789/.0315	1.414/.0323
Coreext/Coreext3sec	14.170/.0988	14.343/**.0978**	**14.113**/.0987	14.430/.0988	14.660/.0988	14.343/.0988
Coreext/Coreext5sec	9.770/.0949	9.783/**.0936**	**9.610**/.0943	9.755/.0959	10.085/.0956	9.812/.0956
Coreext/Coreext10sec	5.672/**.0755**	5.722/.0791	**5.628**/.0784	5.887/.0785	6.147/.0776	5.830/.0777
Coreext/Coreext20sec	3.319/.0592	**3.27**/.0598	3.377/.0612	3.282/.0606	3.603/.0616	3.280/**.0590**
Coreext/Coreext40sec	2.424/**.0451**	2.403/.046	2.612/.0466	2.444/.0452	2.751/.0462	**2.395**/.0460

- The mean m and standard deviation v of training data are first calculated.
- All data including training, enrollment and test data will be Gaussianized. In other words, for every i-Vector w, the new vector will be modified to $w = \dfrac{w-m}{v}$.
- The Gaussianized i-Vectors are length normalized.
- The LDA projection matrix A trained over training data is calculated next.
- All data are then projected into the new feature space. In other words, for every i-Vector w, the transformed i-Vector will be $w = A^T w$.
- The new i-Vectors are length normalized again.
- Finally, these data are used in calculating the cosine kernel defined in Eq. 13.

The linear kernel in GDA is equivalent to the traditional LDA method, and will be compared to the above-mentioned variations of cosine kernel. For training GDA and LDA, a smaller subset of training speech segments (e.g., 10000 vs. 13605) was used (to limit the usage of memory); while, PLDA is trained on the entire male data (e.g., 13605).

4.3. Experimental results

This subsection provides evaluation of speaker recognition comparing the effectiveness of LDA and GDA methods for discriminant analysis and dimensionality reduction.

To assess the system, we use Equal Error Rate (EER) and minimum of decision cost function (minDCF) calculated by,

$$C_{Det} = C_{Miss} \times P_{Miss|Target} \times P_{Target}$$
$$+ C_{FalseAlarm} \times P_{FalseAlarm|NonTarget} \times (1 - P_{Target})$$

Default values for parameters in this equation has been set as $C_{Miss} = C_{FalseAlarm} = 1$ and $P_{Target} = 1/1000$ from the NIST SRE2010 challenge.

Results in terms of EER and minDCF are summarized in Table 2. The EER results show that in all enrollment/test segment condition combinations, GDA improves LDA. Specially, Cosine and WCCN-Cosine have the best results for EER. For minDCF, just in coreext/coreext, coreext/coreext10sec, coreext/coreext40sec combinations, the linear kernel provides slightly better results, but in other cases GDA performed better.

Generally, the improvement of GDA over LDA is more clear in longer duration test segments. Because, i-Vectors extracted for shorter test data are not as accurate as the longer ones; therefore, non-linear discrimination cannot perfectly locate them in their correct speaker classes.

Here, the cosine kernel performs the best among other kernel functions including the linear kernel (or LDA). After that, WCCN-Cosine and Gaussianized-Cosine and Linear kernels achieved effective performance. However, the LDA-Cosine kernel and LDA-WCCN-Cosine unexpectedly did not provide sufficient improvement over the other kernel functions. In summary, experimental results show that GDA is a promising dimensionality reduction and discrimination approach for i-Vector/PLDA system. With GDA the relative improvement of 20% in EER and 18% in minDCF with core/core condition is achieved.

5. Conclusion and future work

In this paper, we have studied the usefulness of GDA in state-of-the-art i-Vector based speaker recognition using PLDA scoring. Most speaker recognition approaches use LDA to separate speaker classes and reduce the dimensionality of the feature vectors. Alternatively, GDA relaxes the linear separability of classes, which can be effective if unknown distortion or mismatch is present. We used NIST SRE2010 core and coreext conditions for experiments, and results show that GDA achieves effective gains for improving i-Vector/PLDA systems.

The cosine kernel, using WCCN before cosine kernel, and the Gaussianized cosine kernel achieved better performance compared to other kernel functions. The combination of LDA and WCCN before the cosine kernel did not provide improvement. The LDA used here merely separates speaker classes and does not perform dimensionality reduction. In the future, a cascade of LDA and WCCN with dimensionality reduction to various sizes will be studied in combination with the cosine kernel.

The relative gain for short test segments is not as comparable as that for the original long versions. Therefore, for future work, we would like to introduce uncertainty of i-Vectors into GDA to make the system more robust. In previous works, researchers had applied uncertainty of short test segment in LDA, PLDA, or in the scoring phase to obtain improvement. Therefore, GDA using uncertainty information could be better for short test data. In addition, the effectiveness of GDA for i-Vector/PLDA systems when input data contains different ratios of SNR will be studied.

6. Acknowledgment

The authors would like to thank Dr. Soheil Khorram of the University of Michigan for very helpful discussions and comments.

7. References

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

[2] S. J. Prince and J. H. Elder, "Probabilistic linear discriminant analysis for inferences about identity," in *IEEE 11th International Conference on Computer Vision (ICCV)*, 2007, pp. 1–8.

[3] J. H. L. Hansen and T. Hasan, "Speaker recognition by machines and humans: A tutorial review," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 74–99, 2015.

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

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

[6] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[7] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Computation*, vol. 12, no. 10, pp. 2385–2404, 2000.

[8] N. Kumar and A. G. Andreou, "Heteroscedastic discriminant analysis and reduced rank hmms for improved speech recognition," *Speech Communication*, vol. 26, no. 4, pp. 283–297, 1998.

[9] B. Scholkopft and K.-R. Mullert, "Fisher discriminant analysis with kernels," *Neural Networks for Signal Processing IX*, vol. 1, no. 1, p. 1, 1999.

[10] T. Hastie and R. Tibshirani, "Discriminant analysis by gaussian mixtures," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 155–176, 1996.

[11] N. Dehak, P. Kenny, R. Dehak, O. Glembek, P. Dumouchel, L. Burget, V. Hubeika, and F. Castaldo, "Support vector machines and joint factor analysis for speaker verification," in *Proceedings of IEEE ICASSP*, 2009, pp. 4237–4240.

[12] N. Dehak, *Discriminative and generative approaches for long- and short-term speaker characteristics modeling: application to speaker verification*. Ecole de Technologie Superieure (Canada), 2009.

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

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

[15] S. O. Sadjadi, J. W. Pelecanos, and W. Zhu, "Nearest neighbor discriminant analysis for robust speaker recognition." in *Proceedings of INTERSPEECH*, 2014, pp. 1860–1864.

[16] R. Saeidi, R. Astudillo, and D. Kolossa, "Uncertain LDA: Including observation uncertainties in discriminative transforms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, no. 99, 2015.

[17] C. Yu, G. Liu, S. Hahm, and J. H. L. Hansen, "Uncertainty propagation in front end factor analysis for noise robust speaker recognition," in *Proceedings of IEEE ICASSP*, 2014, pp. 4017–4021.

[18] T. Hasan, R. Saeidi, J. H. L. Hansen, and D. A. van Leeuwen, "Duration mismatch compensation for i-vector based speaker recognition systems," in *Proceedings of IEEE ICASSP*, 2013, pp. 7663–7667.

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

[20] "The nist year 2010 speaker recognition evaluation plan," 2010. [Online]. Available: http://www.nist.gov/itl/iad/mig/upload/NIST_SRE10_evalplan-r6.pdf

[21] 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. 345–354, 2005.

[22] G. Liu and J. H. L. Hansen, "An investigation into back-end advancements for speaker recognition in multi-session and noisy enrollment scenarios," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 12, pp. 1978–1992, 2014.