Odyssey 2012
The Speaker and Language Recognition Workshop
25-28 June 2012, Singapore

PLDA based Speaker Verification with Weighted LDA Techniques

Ahilan Kanagasundaram, David Dean, Sridha Sridharan, Robbie Vogt

Speech and Audio Research Laboratory
Queensland University of Technology, Brisbane, Australia
{a.kanagasundaram, d.dean, s.sridharan, r.vogt}@qut.edu.au

Abstract

This paper investigates the use of the dimensionality-reduction techniques weighted linear discriminant analysis (WLDA), and weighted median fisher discriminant analysis (WMFD), before probabilistic linear discriminant analysis (PLDA) modeling for the purpose of improving speaker verification performance in the presence of high inter-session variability. Recently it was shown that WLDA techniques can provide improvement over traditional linear discriminant analysis (LDA) for channel compensation in i-vector based speaker verification systems. We show in this paper that the speaker discriminative information that is available in the distance between pair of speakers clustered in the development i-vector space can also be exploited in heavy-tailed PLDA modeling by using the weighted discriminant approaches prior to PLDA modeling. Based upon the results presented within this paper using the NIST 2008 Speaker Recognition Evaluation dataset, we believe that WLDA and WMFD projections before PLDA modeling can provide an improved approach when compared to uncompensated PLDA modeling for i-vector based speaker verification systems.

1. Introduction

I-vector-based speaker verification has recently become the state of the art of speaker verification, providing superior performance when compared to joint factor analysis (JFA) approach [1]. Rather than taking the JFA approach of modeling speaker and channel variability spaces separately, the i-vector approach forms a low-dimensional, total-variability space that models both speaker and channel variability together. Unlike JFA, where factor analysis is used to generate a discriminative model, the i-vector approach uses similar factor analysis techniques as a feature extractor, creating an intermediate speaker representation between the high dimensional Gaussian mixture model (GMM) super-vector and traditional low dimensional acoustic feature representations [1]. As the channel variation is included within the total variability space, i-vector features are often combined with channel compensation techniques to attenuate channel variation in the i-vector space. The choice of channel compensation techniques have become a very active area of research, with initial research focusing on the use of linear discriminant analysis (LDA) followed by within-class covariance normalization (WCCN), as proposed by Dehak *et al.* [2]. Recently, this approach was extended by McLaren and van Leeuwen [3] who proposed a new LDA-based approach, source-normalized LDA (SN-LDA), which improves the i-vector speaker representation in both mismatched conditions and conditions for which limited hyperparameter developmental speech resources are available. This work has been futher extended by Kanagasundaram *et al.*, by investigating new channel compensation approaches of weighted LDA (WLDA) and source-normalized weighted LDA (SN-WLDA) [4], and these were found to achieve further improvement over both the non-weighted LDA and SN-LDA techniques.

Recently these low dimensional i-vector features were extended with a probabilistic linear discriminant analysis (PLDA) approach to model speaker and channel part within the i-vector space, and this has been shown to provide improved speaker verification performance to the initial i-vector approach [5, 6, 7]. This PLDA technique was originally proposed by Price *et al.* [8] for face recognition, and was adapted to i-vectors for speaker verification by Kenny *et al.* [5, 6, 7]. In his original paper, Kenny investigated two generative approaches to forming the PLDA models: Gaussian PLDA (GPLDA) and heavy-tailed PLDA (HTPLDA) [5]. Kenny found that HTPLDA achieved significant improvement over GPLDA, concluding that i-vector features are better modeled by heavy-tailed distribution due to the frequent presence of outliers in the i-vector space. More recently Matejka *et al.* have investigated dimensionality reduction using LDA before PLDA modeling [9], and achieved an improvement on *telephone-telephone* (*enrolment-verification*) conditions. However this approach of transforming the i-vector space before PLDA modeling has not yet been investigated under mismatched conditions where *enrolment* and *verification* conditions are not matched. More importantly, the investigation of more advanced channel compensation techniques such as WLDA, median fisher discriminator (MFD), and weighted MFD (WMFD) would be of considerable value to improving PLDA-based speaker verification systems.

The advantages of LDA-based approaches is that a higher dimensional i-vector feature can be projected into a much lower dimensional space with minimal loss of discriminantive ability, as the ratio of between-speaker and within-speaker variations is maximized. The between-speaker variation normally depends on speaker's characteristics, but the within-speaker variation is much more dependent on the choice of microphone, the acoustic environment, transmission channels and day-to-day differences within a speakers voice. The full potential of using LDA-based approaches with i-vector speaker verification system is not realized with traditional LDA due to the large channel variation and the heavy-tailed behavior of i-vector distributions. We investigate in this paper if channel compensation using LDA, WLDA, MFD, and WMFD can provide superior performance for HTPLDA based speaker verification over non-channel compensated approaches.

This paper is structured as follows: Section 2 gives a brief introduction to the process of PLDA based speaker verification and also introduces i-vector feature extraction, dimensionality reduction techniques, PLDA modeling and scoring. Section 3 describes the methodology of the experiments conducted in this paper, and results and corresponding discussions are given in Section 4. Section 5 concludes the paper.

2. Speaker verification using PLDA techniques

2.1. I-vector feature extraction

I-vectors represent a speaker and channel-specific GMM super-vector by a single total-variability space. This single-subspace approach was motivated by the discovery that the channel space of JFA contains information that can be used to distinguish between speakers [10]. An i-vector speaker and channel dependent GMM super-vector can be represented by,

$$\boldsymbol{\mu} \quad = \quad \mathbf{m} + \mathbf{T}\mathbf{w}, \tag{1}$$

where $\mathbf{m}$ is the same universal background model (UBM) super-vector used in the JFA approach and $\mathbf{T}$ is a low rank total variability matrix. The total-variability factors, $\mathbf{w}$, has a standard normal distribution *N(0,1)*, and is referred to as *i-vectors*. Extracting an i-vector from the total variability subspace is essentially a *maximum a-posteriori adaptation* (MAP) of $\mathbf{w}$ in the subspace defined by $\mathbf{T}$. An efficient procedure for the optimization of the total variability subspace $\mathbf{T}$ and subsequent extraction of i-vectors is described in [11] and [2].

The total variability subspace is responsible for defining a suitable space from which i-vectors are extracted. A pooled total variability approach is used for i-vector feature extraction in this paper, where the total variability subspace (${R_w}^{telmic} = 500$) is trained on telephone and microphone speech pooled utterances. This approach has been found by the authors to provide an improvement over the traditional concatenated approach to multi-condition factor analysis, an analysis of which will be published by the authors separately in the near future.

2.2. Dimensionality reduction of i-vector features

2.2.1. LDA and weighted LDA

In the existing literature, a sequential approach of following LDA by WCCN (LDA + WCCN) has proved useful for speaker verification [1], and an extention of this approach using WLDA, WLDA + WCCN [4], has provided further improvements. In the first stage of the LDA + WCCN sequential approach, LDA is used to define a new spatial axes $\boldsymbol{A}$ that minimizes the within-class variance caused by channel effects and maximizes the variance between speakers in the i-vector space. WCCN is then used as an additional channel compensation technique to scale the subspace in order to attenuate dimensions of high within-class variance.

Both LDA and WCCN calculations are based up the standard within- and between-class scatter estimations S_b and S_w, calculated as

$$\mathbf{S}_b \quad = \quad \sum_{s=1}^{S} n_s(\bar{\mathbf{w}}_s - \bar{\mathbf{w}})(\bar{\mathbf{w}}_s - \bar{\mathbf{w}})^T, \tag{2}$$

$$\mathbf{S}_w \quad = \quad \sum_{s=1}^{S}\sum_{i=1}^{n_s}(\mathbf{w}_i^s - \bar{\mathbf{w}}_s)(\mathbf{w}_i^s - \bar{\mathbf{w}}_s)^T, \tag{3}$$

where S is the total number of speakers, n_s is number of utterances of speaker s. The mean i-vectors, $\bar{\mathbf{w}}_s$ for each speaker, and $\bar{\mathbf{w}}$ is the across all speakers are defined by

$$\bar{\mathbf{w}}_s = \frac{1}{n_s}\sum_{i=1}^{n_s}\mathbf{w}_i^s, \tag{4}$$

$$\bar{\mathbf{w}} = \frac{1}{N}\sum_{s=1}^{S}\sum_{i=1}^{n_s}\mathbf{w}_i^s. \tag{5}$$

where N is the total number of sessions. In the first stage, LDA attempts to find a reduced set of axes $\mathbf{A}$ that minimizes the within-class variability while maximizing the between-class variability through the eigenvalue decomposition of $\mathbf{S}_b\mathbf{v} = \lambda\mathbf{S}_w\mathbf{v}$.

In the second stage, the WCCN transformation matrix ($\boldsymbol{B}$) is trained using the LDA-projected i-vectors from the first stage. The WCCN matrix ($\boldsymbol{B}$) is calculated using Cholesky decomposition of $\boldsymbol{B}\boldsymbol{B}^T = \boldsymbol{W}^{-1}$, where the within-class covariance matrix $\boldsymbol{W}$ is calculated using

$$\mathbf{W} \quad = \quad \frac{1}{S}\sum_{s=1}^{S}\sum_{i=1}^{n_s}(\mathbf{A}^T(\mathbf{w}_i^s - \bar{\mathbf{w}}_s))(\mathbf{A}^T(\mathbf{w}_i^s - \bar{\mathbf{w}}_s))^T. \tag{6}$$

Traditional LDA approach attempts to project high dimensional i-vectors into a more discriminative lower-dimensional subspace. However, this approach does not take advantage of the discriminative relationships that can be found between pairs of classes. This is particularly the case when pairs are positioned closely together, often due to channel similarities, and traditional estimation of the between-class scatter matrix are not able to adequately compensate. The WLDA technique has been used to overcome these shortcoming [4], by refining the between-class scatter matrix through the addition of a weighting function, $w(d_{ij})$, calculated according to the between-class distance of each pair of classes i and j. This weighted between-class scatter matrix, $\mathbf{S}_b^w$, is defined as

$$\mathbf{S}_b^w \quad = \quad \frac{1}{N}\sum_{i=1}^{S-1}\sum_{j=i+1}^{S} w(d_{ij})n_i n_j(\bar{\mathbf{w}}_i - \bar{\mathbf{w}}_j)(\bar{\mathbf{w}}_i - \bar{\mathbf{w}}_j)^T \tag{7}$$

where $\bar{\mathbf{w}}_x$, and n_x is the mean i-vector and session count respectively of speaker x.

In equation (7), the weighting function $w(d_{ij})$ is defined such that the classes that are closer to each other will be more heavily weighted. As the authors have previously shown in [4], when $w(d_{ij})$ equals 1, the weighted between-class scatter estimations will converge to the standard non-weighted between-class scatter form as described in equation (2). In this paper, we will be investigating two weighting functions based on the Euclidean distance, and the Mahalanobis distance between the pairs.

The Euclidean distance weighting function, ${w_{(d_{ij})}}^{Euclidean}$, and the Mahalanobis distance weighting function ${w_{(d_{ij})}}^{Mahalanobis}$, can be defined as follows,

$${w_{(d_{ij})}}^{Euclidean} \quad = \quad ((\bar{\mathbf{w}}_i - \bar{\mathbf{w}}_j)^T(\bar{\mathbf{w}}_i - \bar{\mathbf{w}}_j))^{-n} \tag{8}$$

$${w_{(d_{ij})}}^{Mahalanobis} \quad = \quad ((\bar{\mathbf{w}}_i - \bar{\mathbf{w}}_j)^T(\mathbf{S}_w)^{-1}(\bar{\mathbf{w}}_i - \bar{\mathbf{w}}_j))^{-n} \tag{9}$$

where $\bar{\mathbf{w}}_i$ and $\bar{\mathbf{w}}_j$ are the mean i-vectors of speaker i and j respectively, and the within-class scatter matrix, $\mathbf{S}_w$, is defined by equation (3). In this paper, classification performance will be analyzed with several arbitrary values of n.

The Euclidean distance weighting function, is a monotonically-decreasing function, so neighboring classes closer together will be heavily weighted than neighboring classes wider.

The Mahalanobis distance based weighting function provides some advantages for i-vector speaker representations. If the session i-vectors ($\mathbf{w}_i^s$) are uncorrelated in each speaker and are scaled so that they had unit variances, then $\mathbf{S}_w$ would be the identity matrix and Mahalanobis distance will converge as Euclidean distance between $\bar{\mathbf{w}}_i$ and $\bar{\mathbf{w}}_j$. But there is some correlation between session i-vectors in each speaker and within-class

scatter is not an identity matrix. It can be seen that the presence of within-class scatter ($\mathbf{S}_w$) in the quadratic form (9) will allow for the different scales on which the variables are measured and for non-zero correlations between the variables. If two between-speaker distributions are very close, the Mahalanobis function increases the distance between the classes further than the Euclidean function reflecting less overlap between those classes.

Once the weighted between-class scatter, $\mathbf{S}_b^w$, is estimated for the chosen weighting function, the standard within-class scatter S_w and the corresponding WLDA and WCCN transformation matrices can be estimated as described in section 2.2.1.

2.2.2. MFD and Weighted MFD

In traditional LDA, the speaker-mean i-vector plays a central role in the definition of the between-class and within-class scatter matrices. Therefore the accuracy of its estimate will have a substantial effect on the resulting projection directions of the LDA transformation. In a typical speaker verification systems, each individual will only provide a few recording for training, and averaging these training recording often leads to loss of useful speaker-discriminant information. In this section, MFD and WMFD are investigated to attempt to attenuate this loss. Like the sample average, the median can also be used as an estimator for the central tendency, moreover, it is generally considered that the median is a more robust estimator of the central tendency than the sample average for data with outliers [12]. MFD estimation is based up the median based between- and with-class scatter estimations, $\boldsymbol{S}_w^{median}$ and $\boldsymbol{S}_b^{median}$, calculated as in (2) and (3);

$$\mathbf{S}_b^{median} = \sum_{s=1}^{S} n_s(\bar{\mathbf{w}}_s - \bar{\mathbf{w}})(\bar{\mathbf{w}}_s - \bar{\mathbf{w}})^T, \quad (10)$$

$$\mathbf{S}_w^{median} = \sum_{s=1}^{S}\sum_{i=1}^{n_s}(\mathbf{w}_i^s - \bar{\mathbf{w}}_s)(\mathbf{w}_i^s - \bar{\mathbf{w}}_s)^T \quad (11)$$

where S is the total number of speakers, n_s is number of utterances of speaker s. The median i-vectors, $\bar{\mathbf{w}}_s$ for each speaker, and $\bar{\mathbf{w}}$ across all speakers are defined by

$$\bar{\mathbf{w}}_s = \textit{Median}(\{\mathbf{w}_1^s, \mathbf{w}_2^s, \mathbf{w}_3^s ... \mathbf{w}_{n_s}^s\}), \quad (12)$$

$$\bar{\mathbf{w}} = \frac{1}{N}\sum_{s=1}^{S} n_s \bar{\mathbf{w}}_s. \quad (13)$$

where N is the total number of sessions. The median based weighted between-class matrix ($\mathbf{S}_b^{w-median}$) is estimated using equation (7), where the mean i-vectors of a speaker are replaced with the median instead of mean. Once the median between- and within-class estimations are calculated using equations (10) and (11), MFD and WMFD can be estimated using similar approaches to LDA-based eigenvector decomposition.

2.2.3. I-vector projection with dimensionality reduction techniques

Channel compensation techniques such as LDA, WCCN, WLDA, MFD, and WMFD have been described in sections 2.2.1 and 2.2.2. LDA followed by WCCN channel compensated i-vector can be calculated as follows,

$$\hat{\mathbf{w}} = \boldsymbol{B}^T\boldsymbol{A}^T\mathbf{w} \quad (14)$$

where the estimation of LDA ($\boldsymbol{A}$) and WCCN ($\boldsymbol{B}$) projection matrices have been described in Section 2.2. Channel compensated i-vector, $\hat{\mathbf{w}}$, will be used for heavy-tailed PLDA modeling as explained in Section 2.3.

2.3. HTPLDA modeling

Rather than attempting to model speaker and channel variability in the i-vector space only, a more sophisticated attempt is to model the two variability factors directly in the channel compensated i-vector space. A speaker and channel dependent i-vector, $\hat{\mathbf{w}}$, can be defined as

$$\hat{\mathbf{w}}_r = \bar{\mathbf{w}} + \mathbf{U}_1\mathbf{x}_1 + \mathbf{U}_2\mathbf{x}_{2r} + \boldsymbol{\varepsilon}_r \quad (15)$$

where for given speaker recordings $r = 1, ...R$; $\mathbf{U}_1$ is the eigenvoice matrix and $\mathbf{U}_2$ is the eigenchannel matrix, $\mathbf{x}_1$ and $\mathbf{x}_{2r}$ are the speaker and channel factors respectively and $\boldsymbol{\varepsilon}_r$ is the residuals. In the PLDA modeling approach, the speaker specific part can be represented as $\bar{\mathbf{w}} + \mathbf{U}_1\mathbf{x}_1$, which represents the between speaker variability. The covariance matrix of the speaker part is $\mathbf{U}_1{\mathbf{U}_1}^T$. The channel specific part can be represented as $\mathbf{U}_2\mathbf{x}_{2r} + \boldsymbol{\varepsilon}_r$, which describes the within speaker variability. The covariance matrix of channel part is $\boldsymbol{\Lambda}^{-1} + \mathbf{U}_2{\mathbf{U}_2}^T$. We assume that precision matrix ($\boldsymbol{\Lambda}$) is full rank and remove the eigenchannels ($\mathbf{U}_2$) from equation (15). This is done because the PLDA speaker verification approach didn't show major improvement with eigenchannels, and removing them provides a benefit in reduced computational complexity.

For HTPLDA, Kenny proposed Student's t-distribution as an alternative to the Gaussian for modeling the speaker and channel subspaces in the i-vector space [5]. In HTPLDA, it is assumed that speaker factors and residual factors have heavy-tailed distribution, scaled by gamma distribution scalars, which can be represented as follows,

$$\mathbf{x}_1 \sim N(0, u_1^{-1}I) \text{ where } u_1 \sim G(n_1/2, n_1/2)$$
$$\boldsymbol{\varepsilon}_r \sim N(0, \upsilon_r^{-1}\boldsymbol{\Lambda}^{-1}) \text{ where } \upsilon_1 \sim G(\nu/2, \nu/2)$$

where n_1 and ν are the degrees of freedom, and u_1, υ_r are gamma distribution scalers, $N(\mu, \Sigma)$ represents a Gaussian distribution with mean μ and covariance Σ, and G(a,b) represents a gamma distribution with shape parameter a and scale parameter b. In HTPLDA, the model parameters, $\mathbf{m}$, $\mathbf{U}_1$, $\boldsymbol{\Lambda}$, n_1, and ν are estimated from the development i-vectors.

2.4. PLDA scoring

Scoring in PLDA i-vector speaker verfication systems is conducted using the batch likelihood ratio between a target and test i-vector [5]. Given two i-vectors, $\hat{w}_{target}$ and $\hat{w}_{test}$, the batch likelihood ratio can be calculated as follows,

$$\ln \frac{P(\hat{w}_{target}, \hat{w}_{test} \mid H_1)}{P(\hat{w}_{target} \mid H_0)P(\hat{w}_{test} \mid H_0)} \quad (16)$$

where H_1 denotes the hypothesis that the i-vectors represent the same speakers and H_0 denotes the hypothesis that they do not.

3. Methodology

The PLDA experiments were evaluated using the NIST 2008 Speaker Recognition Evaluation (SRE) utterances from the *short2-short3* evaluation conditions. Performance was evaluated using the equal error rate (EER) and minimum decision cost function (DCF) calculated using $C_{miss} = 10$, $C_{FA} = 1$, and $P_{target} = 0.01$. Evaluation was performed on the NIST

Table 1: *Comparison of baseline systems and HTPLDA systems with LDA and WLDA projections on the common set of the 2008 NIST SRE short2-short3 conditions.*

System	Score norm	Interview-interview		Interview-telephone		Telephone-interview		Telephone-telephone	
		EER	DCF	EER	DCF	EER	DCF	EER	DCF
Baseline system (without dimensionality reduction)									
HTPLDA	Without Norm	6.08%	0.0275	6.36%	0.0286	4.21%	0.0217	**2.55%**	**0.0136**
	With S-Norm	4.50%	0.0230	5.53%	0.0260	3.86%	0.0209	2.88%	0.0173
LDA projected HTPLDA system									
Standard LDA-HTPLDA	Without Norm	6.71%	0.0306	7.17%	0.0322	5.43%	0.0253	3.62%	0.0189
	With S-Norm	4.14%	0.0194	5.82%	0.0245	3.60%	0.0156	2.87%	0.0153
WLDA projected HTPLDA system									
W-LDA-HTPLDA	Without norm	6.16%	0.0288	6.36%	0.0271	4.57%	0.0216	2.65%	0.0151
(w=Euclidean function)	With S-Norm	4.04%	0.0191	**5.44%**	**0.0221**	3.19%	**0.0147**	2.71%	0.0145
W-LDA-HTPLDA	Without norm	5.39%	0.0244	6.08%	0.0274	4.14%	0.0205	**2.55%**	0.0148
(w=Mahalanobis function)	With S-Norm	**3.90%**	**0.0184**	5.53%	**0.0221**	**3.14%**	0.0154	2.63%	0.0147

Table 2: *Comparison of baseline systems and HTPLDA systems with MFD and WMFD projections on the common set of the 2008 NIST SRE short2-short3 conditions.*

System	Score norm	Interview-interview		Interview-telephone		Telephone-interview		Telephone-telephone	
		EER	DCF	EER	DCF	EER	DCF	EER	DCF
Baseline system (without dimensionality reduction)									
HTPLDA	Without S norm	6.08%	0.0275	6.36%	0.0286	4.21%	0.0217	2.55%	**0.0136**
	With S norm	4.50%	0.0230	5.53%	0.0260	3.86%	0.0209	2.88%	0.0173
MFD projected HTPLDA system									
Standard MFD-HTPLDA	Without norm	7.29%	0.0336	7.10%	0.0301	5.30%	0.0261	3.13%	0.0182
	With S-Norm	4.15%	**0.0192**	5.89%	0.0235	3.33%	**0.0146**	2.72%	0.0150
WMFD projected HTPLDA system									
W-MFD-HTPLDA	Without norm	6.54%	0.0302	6.45%	0.0276	4.35%	0.0208	**2.39%**	0.0143
(w=Euclidean function)	With S-Norm	4.12%	0.0201	**5.34%**	**0.0219**	**3.19%**	0.0148	2.63%	0.0143
W-MFD-HTPLDA	Without norm	6.04%	0.0279	6.45%	0.0272	4.48%	0.0207	**2.39%**	0.0149
(w=Mahalanobis function)	With S-Norm	**4.02%**	0.0193	5.53%	0.0220	3.26%	0.0152	2.63%	0.0148

08 DET conditions 3, 4, 5 and 7, corresponding to *interview-interview*, *interview-telephone*, *telephone-interview*, and *telephone-telephone* (English-only) *enrolment-verification* trials [13].

We used 13 feature-warped mel frequency cepstral coefficients (MFCC) with appended delta coefficients and two gender dependent UBMs containing 512-mixture Gaussians throughout our experiments. The UBMs were trained on NIST 2004 SRE corpus, and were used to calculate the Baum-Welch statistics used for further calculation of a total variability subspace of dimension $R_w = 500$. I-vectors were projected into LDA space using 150 eigenvectors.

The pooled total-variability representation and the HTPLDA parameters were trained using telephone and microphone speech data from NIST 2004, 2005 and 2006 SRE corpora as well as Switchboard II. We empirically selected the number of eigenvoices (N_1) equal to 100 as best value according to speaker verification performance. A full precision matrix was used for $\mathbf{\Lambda}$, rather than the diagonal. S-Normalization was applied to telephone and microphone speech based experiments [14], with the statistics calculated using utterances selected from NIST04, 05 and 06 telephone and microphone pooled utterances. The best value of WLDA and WMFD weighting parameter (n) was selected as 4 for Euclidean distance based weighting function and 3 for Mahalanobis distance based weighting function.

4. Results and discussion

Table 1 presents results comparing the performance between the baseline system (without dimensionality reduction) and the HTPLDA system with LDA and WLDA projections on the standard NIST SRE 08 evaluation conditions. These results have show that the WLDA projected HTPLDA system achieved better performance than baseline system (without dimensionality reduction) on all the *enrolment-verification* conditions except *telephone-telephone*. The WLDA projected HTPLDA system achieved over 13% and 20% improvement on EER and DCF respectively for *interview-interview* and *telephone-interview* conditions. However it appears that the LDA and WLDA projected HTPLDA system do not show any improvement in telephone-telephone conditions, and it appears that LDA and WLDA projections change the telephone speech based i-vector distribution.

Table 2 presents results comparing the performance between the baseline system (without dimensionality reduction) and HTPLDA system with MFD and WMFD projections on the standard NIST SRE 08 evaluation standard condition. HTPLDA system with WMFD projection achieved improved performance over the baseline (without dimensionality reduction) and the MFD projected HTPLDA systems on all the *enrolment-verification* conditions. When the WLDA-projected HTPLDA system was compared to the WMFD-projected HTPLDA system, the WMFD-projected HTPLDA system achieved better performance on the *interview-telephone* and *telephone-*

telephone conditions. The WMFD-projected HTPLDA system achieved over 6% improvement when compared to the baseline system for EER in the *telephone-telephone* condition. It has also been shown that there appears to be no big difference in performance between the Euclidean distance weighting function and the Mahalanobis distance weighting function. Results also suggest that S-Normalization improved the performance of LDA-projected HTPLDA systems across all *enrolment-verification* conditions.

5. Conclusion

In this paper, we have investigated LDA, WLDA, MFD, and WMFD projective channel compensation approaches prior to HTPLDA speaker verification. By using the weighted-pairwise Fisher criterion, WLDA and WMFD techniques have been shown to take advantage of the speaker-discriminative information present in the pairwise distances between classes to provide improved speaker verification performance. Through evaluations performed on the NIST 2008 SRE data, both the WLDA and WMFD projected HTPLDA system have shown an improvement in speaker verification performance in both matched and mismatched *enrolment-verification* conditions, with the highest improvement in the *interview-interview* and *telephone-interview enrolment-verification* conditions.

Based upon the results presented within this paper using the NIST 2008 speaker recognition evaluation dataset, we believe that WLDA and WMFD projections before PLDA modeling can be a improved approach when compared to non-channel-compensated PLDA speaker verification systems.

6. Acknowledgements

This project was supported by the Cooperative Research Centre for Advanced Automotive Technologies (AutoCRC).

7. References

[1] 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. PP, no. 99, pp. 1 –1, 2010.

[2] N. Dehak, R. Dehak, J. Glass, D. Reynolds, and P. Kenny, "Cosine similarity scoring without score normalization techniques," *Odyssey Speaker and Language Recognition Workshop*, 2010.

[3] M. McLaren and D. van Leeuwen, "Source-normalised-and-weighted LDA for robust speaker recognition using i-vectors," in *in IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 5456–5459, 2011.

[4] A. Kanagasundaram, D. Dean, R. Vogt, M. McLaren, S. Sridharan, and M. Mason, "Weighted LDA techniques for i-vector based speaker verification," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 2012(accepted).

[5] P. Kenny, "Bayesian speaker verification with heavy tailed priors," in *Proc. Odyssey Speaker and Language Recogntion Workshop, Brno, Czech Republic*, 2010.

[6] M. Senoussaoui, P. Kenny, N. Brummer, E. de Villiers, and P. Dumouchel, "Mixture of PLDA models in i-vector space for gender independent speaker recognition," *Proceed. of INTERSPEECH*, pp. 25–28, 2011.

[7] L. Burget, O. Plchot, S. Cumani, O. Glembek, P. Matejka, and N. Brümmer, "Discriminatively trained probabilistic linear discriminant analysis for speaker verification," pp. 4832–4835, ICASSP, 2011.

[8] J. Price and T. Gee, "Face recognition using direct, weighted linear discriminant analysis and modular subspaces," *Pattern Recognition*, vol. 38, no. 2, pp. 209–219, 2005.

[9] P. Matejka, O. Glembek, F. Castaldo, M. 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*, pp. 4828–4831, IEEE, 2011.

[10] N. Dehak, R. Dehak, P. Kenny, N. Brummer, P. Ouellet, and P. Dumouchel, "Support vector machines versus fast scoring in the low-dimensional total variability space for speaker verification," in *Proceedings of Interspeech*, p. 1559 1562, 2009.

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

[12] J. Yang, J. Yang, and D. Zhang, "Median fisher discriminator: a robust feature extraction method with applications to biometrics," *Frontiers of Computer Science in China*, vol. 2, no. 3, pp. 295–305, 2008.

[13] "The NIST year 2008 speaker recognition evaluation plan," tech. rep., NIST, April 2008.

[14] S. Shum, N. Dehak, R. Dehak, and J. Glass, "Unsupervised speaker adaptation based on the cosine similarity for text-independent speaker verification," *Proc. Odyssey*, 2010.