Available online at www.sciencedirect.com

SciVerse ScienceDirect

Computer Speech and Language 28 (2014) 121–140

COMPUTER SPEECH AND LANGUAGE

www.elsevier.com/locate/csl

I-vector based speaker recognition using advanced channel compensation techniques☆

Ahilan Kanagasundaram[a,*], David Dean[a], Sridha Sridharan[a], Mitchell McLaren[b], Robbie Vogt[a]

[a] *Speech and Audio Research Lab, SAIVT, Queensland University of Technology, Australia*
[b] *Centre for Language and Speech Technology, Radboud University Nijmegen, The Netherlands*

Received 21 September 2012; received in revised form 27 March 2013; accepted 3 April 2013
Available online 25 April 2013

Abstract

This paper investigates advanced channel compensation techniques for the purpose of improving i-vector speaker verification performance in the presence of high intersession variability using the NIST 2008 and 2010 SRE corpora. The performance of four channel compensation techniques: (a) weighted maximum margin criterion (WMMC), (b) source-normalized WMMC (SN-WMMC), (c) weighted linear discriminant analysis (WLDA) and (d) source-normalized WLDA (SN-WLDA) have been investigated. We show that, by extracting the discriminatory information between pairs of speakers as well as capturing the source variation information in the development i-vector space, the SN-WLDA based cosine similarity scoring (CSS) i-vector system is shown to provide over 20% improvement in EER for NIST 2008 interview and microphone verification and over 10% improvement in EER for NIST 2008 telephone verification, when compared to SN-LDA based CSS i-vector system. Further, score-level fusion techniques are analyzed to combine the best channel compensation approaches, to provide over 8% improvement in DCF over the best single approach, SN-WLDA, for NIST 2008 interview/telephone enrolment-verification condition. Finally, we demonstrate that the improvements found in the context of CSS also generalize to state-of-the-art GPLDA with up to 14% relative improvement in EER for NIST SRE 2010 interview and microphone verification and over 7% relative improvement in EER for NIST SRE 2010 telephone verification.

Keywords: Speaker verification; I-vector; GPLDA; LDA; SN-LDA; WLDA; SN-WLDA

1. Introduction

Recent research in speaker verification has focused on the i-vector features based on front-end factor analysis. This technique was originally proposed by Dehak et al. (2010) to provide an intermediate speaker representation between the high-dimensional Gaussian mixture model (GMM) super-vector and traditional low-dimensional mel-frequency cepstral coefficients (MFCCs) feature representation. The extraction of these intermediate-sized vectors, or i-vectors, was motivated by the existing super-vector-based joint factor analysis (JFA) approach (Kenny, 2005; Kenny et al.,

☆ This paper has been recommended for acceptance by Haizhou Li.

* Corresponding author.
E-mail addresses: a.kanagasundaram@qut.edu.au, ahilan.kanagasundaram@student.qut.edu.au (A. Kanagasundaram), d.dean@qut.edu.au (D. Dean), s.sridharan@qut.edu.au (S. Sridharan), m.mclaren@let.ru.nl (M. McLaren), r.vogt@qut.edu.au (R. Vogt).

0885-2308/$ – see front matter
http://dx.doi.org/10.1016/j.csl.2013.04.002

2008). While the JFA approach models the speaker and channel variability space separately, i-vectors are formed by modeling a single low-dimensional total-variability space that covers both the speaker and channel variability (Dehak et al., 2010). This approach was motivated by Dehak et al. finding that i-vectors do not lose any speaker discriminant information, unlike the JFA approach, where some speaker discriminant information is lost in the channel space (Dehak et al., 2010). As the channel variability is included within the total-variability space, Dehak et al. (2010) had investigated a number of standard channel compensation techniques, including linear discriminant analysis (LDA), within-class covariance normalization (WCCN) and nuisance attribute projection (NAP) to attenuate channel variability in the i-vector space.

The i-vector framework was extended with a probabilistic linear discriminant analysis (PLDA) approach to model the speaker and channel parts within the i-vector space, and this has been shown to provide an improved speaker verification performance over cosine similarity scoring (CSS) with channel compensation (Kenny, 2010; Matejka et al., 2011; Senoussaoui et al., 2011). We believe that this is because the uncompensated i-vector behavior is heavy-tailed, and heavy-tailed PLDA (HTPLDA) can explicitly model outliers in the i-vector space (Kenny, 2010). Recently, Garcia-Romero et al. have introduced length-normalized Gaussian PLDA (GPLDA) approach in Garcia-Romero and Espy-Wilson (2011), which has shown similar performance as HTPLDA and it is an approach more computationally efficient than HTPLDA.

Channel variability can be defined as mismatch between enrolment and verification utterances, arising from the differences in microphones, acoustic environment, transmission channels and the variation in individual speaker's voices. Channel compensation can occur at several levels, such as feature domain, model domain and score domain in an i-vector speaker verification system. Feature warping techniques are commonly used in the feature domain, which provides a robustness to additive noise and linear channel mismatch while retaining the speaker specific information (Pelecanos and Sridharan, 2001). In the model domain, an (WCCN[LDA]) approach, which represents the sequential operation of LDA followed by WCCN approach, was used by Dehak et al. (2010) to show good performance. More recently, this approach was extended by McLaren and van Leeuwen (2011) by proposing a new LDA-based approach, source-normalized LDA (SN-LDA), which improves i-vector-based speaker recognition in both mismatched conditions and conditions for which limited system development speech resources are available. In the score domain, t-normalization addresses the problem of session variability by compensating the mismatch between enrolment and verification conditions (Auckenthaler et al., 2000). The model domain channel compensation approaches are presently the most active area of research, as most of the channel variations are captured at the model domain.

The LDA channel compensation technique (including SN-LDA) is based upon the ratio of between-class scatter to within-class scatter, which is used to transform the i-vector space to maximize the between-speaker discriminant information (between-class scatter) while minimizing the within-speaker variability (within-class scatter). The between-speaker scatter depends on speakers' characteristics, while the within-speaker scatter depends largely on microphones, acoustic environments, transmission channels and differences in individual speaker's voices. In the standard LDA approach the influence of between- and within-class information on the transformed space is fixed, as it is calculated based on the ratio of between-class scatter to within-class scatter. Research in the similar field of face recognition has demonstrated, however, that this shortcoming could be overcome by making use of the weighted maximum margin criterion (WMMC) (Cheng et al., 2008; Baker et al., 2009; Hu et al., 2010), in which the objective function is calculated as the difference between the between-class scatter and the weighted within-class scatter. The first aim of this paper is to investigate the WMMC (including SN-WMMC) as an alternative channel compensation approach to LDA (including SN-LDA) for i-vector speaker verification.

Most of the channel compensation techniques take direct advantage of the calculated between- and within-class scatter matrices, and can have problems when classes tend to clump according to characteristics external to class identity. Recently we have investigated the weighted LDA (WLDA) technique (Kanagasundaram et al., 2012), based upon the weighted pair-wise Fisher criteria, that has shown promise in the field of face recognition (Loog et al., 2001; Price and Gee, 2005; Liang et al., 2007), by taking advantage of the discriminatory information between pairs of classes. By applying a weighted parameter to class pairs that weights closer pairs higher, WLDA can provide an improvement in the discriminative ability between classes that would otherwise be difficult to distinguish in an LDA-transformed space. We have presented an introduction to the technique for i-vector speaker verification (Kanagasundaram et al., 2012), no detailed study of the application of this technique to i-vector based speaker verification has been performed in the past. In this paper we will be investigating a range of weighting functions for WLDA (and the related SN-WLDA) technique, which could help to increase the distance between the classes.

Previous studies have shown that the best speaker verification performance for CSS classification of i-vectors can be obtained by first reducing the i-vectors dimensionality through LDA, then weighting the dimensions through WCCN (Dehak et al., 2010), and this approach has shown to still work well for more advanced channel compensation techniques, such as SN-LDA, replacing LDA in this process (McLaren and van Leeuwen, 2011, 2012). Accordingly, throughout this paper, we will take a similar approach and test a range of novel advanced channel compensation techniques for i-vector dimensionality reduction, which will then still be followed by a WCCN-based dimensionality weighting. In addition to this chaining approach to channel compensation, we also believe that better performance could be obtained through score fusion of differently channel compensated i-vector systems running in parallel. Accordingly we will also be investigating score fusion of our best channel compensation techniques in this paper to investigate the complementary nature of these techniques.

Finally, we also hypothesize that if we train the most recent state-of-the-art system, length-normalized GPLDA, on channel compensated i-vector features, it could achieve further improvement as the channel variations can be compensated through the channel compensation approach as well as the length-normalized GPLDA modelling. The best channel compensation approach, which will be found from CSS i-vector system experiments, will be analyzed with length-normalized speaker verification system.

This paper is structured as follows: Section 2 gives a brief introduction to the i-vector based speaker verification system. Section 3 initially details the existing channel compensation techniques and also introduces the novel channel compensation techniques in the latter part. The experimental protocol and corresponding results are given in Sections 4 and 5. Section 6 concludes the paper.

2. I-vector based speaker verification

The i-vector based system initially proposed by Dehak et al. (2010), which has recently become a popular approach for efficient text-independent speaker verification, is based on CSS. Initially speaker utterances are represented by their mixture-occupying based Baum–Welch statistics, calculated using a gender-dependent universal background model (UBM) parameter for each given speech utterance (Kenny et al., 2008). These statistics are used to train a total-variability subspace that can then be used for CSS classification, as outlined in the following sections.

2.1. Total-variability subspace training

I-vectors represent the GMM super-vector by a single total-variability space, which was motivated by the discovery by Dehak et al. (2009) that the channel space of JFA contains speaker information that can be used to also distinguish speakers. A speaker- and channel-dependent GMM super-vector in the i-vector framework can be represented by,

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

where $\mathbf{m}$ is the same UBM super-vector used in the JFA approach and $\mathbf{T}$ is a low rank matrix representing the primary directions across all development data. The total-variability factors, $\mathbf{w}$, have a standard normal distribution $\mathcal{N}(0, 1)$ and are referred to as *i-vectors*. An efficient procedure of total-variability subspace, $\mathbf{T}$, training and subsequent i-vector extraction is described in Kenny et al. (2008) and Dehak et al. (2010).

In this paper, we will be investigating a combined telephone and microphone speaker verification system, and for this approach the total-variability subspace should be trained in a manner that best exploits the useful speaker variability contained in speech acquired from both telephone and microphone sources. McLaren and van Leeuwen (2011) have investigated different types of total-variability representations, such as pooled and concatenation with i-vector systems. For the pooled approach, telephone and microphone utterances are pooled together and for the concatenated approach, individual total-variability subspaces are trained on each source-dependent subset of the training data, then a single subspace is formed through the concatenation of each individual subspaces. McLaren and van Leeuwen found that the pooled approach provided a much better representation for telephone and microphone i-vector speaker verification, and had the additional advantage of being a simpler approach than concatenation (McLaren and van Leeuwen, 2011). In this paper, the pooled total-variability approach will be used for i-vector feature extraction.

2.2. CSS classifier

I-vectors were originally considered as a feature for SVM classification, however, fast scoring approaches using a cosine kernel directly as a classifier were found to provide similar performance to SVMs with a considerable increase in efficiency (Dehak et al., 2009). The CSS classifier operates by comparing the angles between a test i-vector, $\hat{\mathbf{w}}_{\text{test}}$, and a target i-vector $\hat{\mathbf{w}}_{\text{target}}$:

$$\mathrm{S}(\hat{\mathbf{w}}_{\text{target}}, \hat{\mathbf{w}}_{\text{test}}) = \frac{\langle \hat{\mathbf{w}}_{\text{target}}, \hat{\mathbf{w}}_{\text{test}} \rangle}{\|\hat{\mathbf{w}}_{\text{target}}\| \|\hat{\mathbf{w}}_{\text{test}}\|}. \tag{2}$$

2.3. Length-normalized GPLDA classifier

The PLDA technique was originally proposed by Prince and Elder (2007) for face recognition, and later it was introduced to speaker verification to model the actual behavior of i-vector features by Kenny (2010), Senoussaoui et al. (2011), and Burget et al. (2011). In his work, Kenny investigated two PLDA approaches, GPLDA and HTPLDA (Kenny, 2010). He found that HTPLDA shows significant improvement over GPLDA as the distribution of the i-vectors is heavy-tailed. Recently, Garcia-Romero et al. have introduced length-normalized GPLDA approach in Garcia-Romero and Espy-Wilson (2011), and it has shown similar performance as HTPLDA, since the length-normalization approach was used to convert the distribution of the i-vectors from heavy-tailed to Gaussian. The length-normalized GPLDA approach is computationally efficient, so we have chosen to use in this paper. As we focus on advanced channel compensation approaches, the length-normalized GPLDA is modelled on channel compensated i-vector features, $\hat{\mathbf{w}}_r$, which can be defined as

$$\hat{\mathbf{w}}_r = \overline{\mathbf{w}} + \mathbf{U}_1 \mathbf{x}_1 + \boldsymbol{\varepsilon}_r \tag{3}$$

where for given speaker recordings $r = 1, \ldots, R$; $\mathbf{U}_1$ is the eigenvoice matrix, $\mathbf{x}_1$ is the speaker factor and $\boldsymbol{\varepsilon}_r$ is the residuals. The between-speaker variability in the PLDA model is represented by the low rank $\mathbf{U}_1\mathbf{U}_1{}^T$ matrix. The within-speaker variability is described by $\boldsymbol{\Lambda}^{-1}$. We assume that precision matrix ($\boldsymbol{\Lambda}$) is full rank.

The details of length-normalization approach and the estimation of model parameters are given in Kenny (2010) and Garcia-Romero and Espy-Wilson (2011). GPLDA based i-vector system scoring calculated using batch likelihood ratio (Kenny, 2010). Batch likelihood calculation is computationally more expensive than CSS. Given two i-vectors $\hat{\mathbf{w}}_{\text{target}}$ and $\hat{\mathbf{w}}_{\text{test}}$, batch likelihood ratio can be calculated as follows,

$$\ln \frac{P(\hat{\mathbf{w}}_{\text{target}}, \hat{\mathbf{w}}_{\text{test}} | H_1)}{P(\hat{\mathbf{w}}_{\text{target}} | H_0) P(\hat{\mathbf{w}}_{\text{test}} | H_0)} \tag{4}$$

where H_1: the speakers are same and H_0: the speaker are different.

3. Channel compensation techniques

In CSS based i-vector system, as i-vectors are defined by a single variability space, containing both speaker and channel information, there is a requirement that additional intersession, or channel compensation approaches be taken before verification. While approaches, such as LDA achieve dimension reduction, our aim is to compensate for the channel variability (McLaren and van Leeuwen, 2012). Channel compensation approaches are estimated based on within- and between-class scatter variances. The within-class scatter depends on microphones, acoustic environments, transmission channels and differences in individual speaker's voices. On the other hand the between-class scatter depends on speaker's characteristics. These channel compensation techniques are typically designed to maximize the effect of between-class variability and minimize the effects of within-class variability. Our main aim of this paper is to identify the best channel compensation approach for telephone and microphone based i-vector speaker verification systems.

3.1. Within class covariance normalization (WCCN)

WCCN is used as a channel compensation technique to scale a subspace in order to attenuate dimensions of high within-class variance. For use in speaker verification, a within-class variance matrix, $\mathbf{W}$, is calculated using

$$\mathbf{W} = \frac{1}{S}\sum_{s=1}^{S}\sum_{i=1}^{n_s}(\mathbf{w}_i^s - \overline{\mathbf{w}}_s)(\mathbf{w}_i^s - \overline{\mathbf{w}}_s)^T, \tag{5}$$

where $\mathbf{w}_i^s$ is the i-vector representation of i session of speaker s, the mean i-vector for each speaker ($\overline{\mathbf{w}}_s$) is equal to $\frac{1}{n_s}\sum_{i=1}^{n_s}\mathbf{w}_i^s$, S is the total number of speakers and n_s is the number of utterances of speaker s. The WCCN matrix, $\mathbf{B}_1$, can be calculated using Cholesky decomposition of $\mathbf{B}_1\mathbf{B}_1{}^T = \mathbf{W}^{-1}$.

The WCCN channel compensated i-vector ($\hat{\mathbf{w}}_{WCCN}$) can be calculated as follows:

$$\hat{\mathbf{w}}_{WCCN} = \mathbf{B}_1{}^T\mathbf{w} \tag{6}$$

3.2. Linear discriminant analysis (LDA)

LDA is used as channel compensation technique, which 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

$$\mathbf{S}_b\mathbf{v} = \lambda\mathbf{S}_w\mathbf{v}. \tag{7}$$

where the between-class, $\mathbf{S}_b$, and within-class scatter, $\mathbf{S}_w$, can be calculated as follows:

$$\mathbf{S}_b = \sum_{s=1}^{S} n_s(\overline{\mathbf{w}}_s - \overline{\mathbf{w}})(\overline{\mathbf{w}}_s - \overline{\mathbf{w}})^T, \tag{8}$$

$$\mathbf{S}_w = \sum_{s=1}^{S}\sum_{i=1}^{n_s}(\mathbf{w}_i^s - \overline{\mathbf{w}}_s)(\mathbf{w}_i^s - \overline{\mathbf{w}}_s)^T, \tag{9}$$

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

$$\overline{\mathbf{w}}_s = \frac{1}{n_s}\sum_{i=1}^{n_s}\mathbf{w}_i^s, \tag{10}$$

$$\overline{\mathbf{w}} = \frac{1}{N}\sum_{s=1}^{S}\sum_{i=1}^{n_s}\mathbf{w}_i^s, \tag{11}$$

where N is the total number of sessions.

The LDA channel compensated i-vector ($\hat{\mathbf{w}}_{LDA}$) can be then calculated as follows:

$$\hat{\mathbf{w}}_{LDA} = \mathbf{A}^T\mathbf{w} \tag{12}$$

3.3. Source-normalized LDA (SN-LDA)

McLaren and van Leeuwen (2011, 2012) found that the between-class scatter calculated using the standard LDA approach can be influenced by source variation under mismatched conditions, where sources were defined as speech recorded using either microphone or telephone. This influence can be reduced by estimating the between-class scatter using source-normalized i-vectors and fixing the within-class scatter as the residual variations in the i-vector space (McLaren and van Leeuwen, 2011). The source-normalized between-class scatter, $\mathbf{S}_b^{src}$, can be composed of the source-dependent between-class scatter matrices for telephone and microphone-recorded speech, which can be calculated as follows:

$$\mathbf{S}_b^{src} = \mathbf{S}_b^{tel} + \mathbf{S}_b^{mic} \tag{13}$$

where

$$\mathbf{S}_b^{tel} = \sum_{s=1}^{S_{tel}} n_s(\overline{\mathbf{w}}_s - \overline{\mathbf{w}}_{tel})(\overline{\mathbf{w}}_s - \overline{\mathbf{w}}_{tel})^T, \tag{14}$$

$$\mathbf{S}_b^{mic} = \sum_{s=1}^{S_{mic}} n_s(\overline{\mathbf{w}}_s - \overline{\mathbf{w}}_{mic})(\overline{\mathbf{w}}_s - \overline{\mathbf{w}}_{mic})^T, \tag{15}$$

where the mean i-vector for telephone source ($\overline{\mathbf{w}}_{tel}$) is equal to $\frac{1}{n_{tel}}\sum_{i=1}^{n_{tel}} \mathbf{w}_i^{tel}$, and the mean i-vector for microphone source ($\overline{\mathbf{w}}_{mic}$) is equal to $\frac{1}{n_{mic}}\sum_{i=1}^{n_{mic}} \mathbf{w}_i^{mic}$. Rather than estimating the within-class scatter separately as in Eq. (9), McLaren et al. calculated the within-class scatter matrix as the difference between a total variance matrix, $\mathbf{S}_t$, and the source-normalized between-class scatter as:

$$\mathbf{S}_w = \mathbf{S}_t - \mathbf{S}_b^{src}, \tag{16}$$

where

$$\mathbf{S}_t = \sum_{n=1}^{N} (\mathbf{w}_n - \overline{\mathbf{w}})(\mathbf{w}_n - \overline{\mathbf{w}})^T. \tag{17}$$

This approach allows $\mathbf{S}_w$ to be more accurately estimated when development dataset does not provide examples of each speech source from every speaker. Similarly to the LDA approach outlined previously, the SN-LDA channel compensated i-vector will be calculated using Eq. (12).

3.4. Weighted maximum margin criterion (WMMC)

In the LDA or SN-LDA approach, the transformation matrix is based on the ratio of between-class scatter to within-class scatter, and the level of importance of within- and between-class scatters cannot be changed. Research in the field of face recognition has found that weighted maximum margin criterion (WMMC) estimation can be used to change the level of importance of within- and between-class scatters by using weighting coefficients (Cheng et al., 2008; Baker et al., 2009; Hu et al., 2010). We will be applying similar techniques with i-vectors to see how performance varies with different level of within- and between-class scatters.

The objective function of WMMC under projection matrix $\mathbf{A}$ is defined as,

$$J(\mathbf{A}) = tr\{\mathbf{A}^T(W \times \mathbf{S}_w - \mathbf{S}_b)\mathbf{A}\}, \tag{18}$$

where an $\mathbf{A}$ that maximizes Eq. (18) can be calculated through the following eigenvalue equation:

$$(W \times \mathbf{S}_w - \mathbf{S}_b)\mathbf{v} = \lambda \mathbf{v}, \tag{19}$$

where the within-class scatter ($\mathbf{S}_w$) and between-class scatter ($\mathbf{S}_b$) are estimated as described in Eqs. (9) and (8). W is a weighting coefficient defining the relative influence of the $\mathbf{S}_w$ and $\mathbf{S}_b$.

In this paper we will be investigating manual weighting coefficients, where performance of WMMC is directly dependent on its weighted coefficient. The WMMC channel compensated i-vector will be calculated using Eq. (12).

We have detailed the SN-LDA approach in Section 3.3, which was previously proposed by McLaren et al. to i-vector system. From the basics of SN-LDA approach, we introduce the SN-WMMC approach to i-vector system, which can be used to improve the performance in both mismatched enrolment/verification conditions. In this case, the between-class scatter matrix ($\mathbf{S}_b$), and within-class scatter matrix ($\mathbf{S}_w$) are estimated using Eqs. (13) and (9).

3.5. Weighted LDA (WLDA)

Traditional LDA techniques attempt to project i-vectors into a more discriminative lower-dimensional subspace, calculated based on within- and between-class scatter matrix estimations. However, this approach cannot take advantage of the discriminative relationships between the class pairs, which are much closer due to channel similarities, and traditional estimation of between-class scatter matrix is not able to adequately compensate. Weighted LDA (WLDA)

technique can be used to overcome this problem (Loog et al., 2001), by weighting the classes that are closer to each other to reduce class confusion. Even though WLDA techniques have been introduced to face recognition recently (Loog et al., 2001), the effective weighting function hasn't been found, which could help to extract more discriminative information. In this paper, we introduce the WLDA approach to i-vector speaker verification and explore the application of several alternative weighting functions to extract more speaker discriminative information. In the WLDA approach, the between-class scatter matrix is redefined by adding a weighting function, $w(d_{ij})$, according to the between-class distance of each pair of classes i and j. In Loog et al. (2001), the equations, which are used to calculate the within- and between-class scatter estimations, are bit different from equations that are used in i-vector speaker verification (McLaren and van Leeuwen, 2011; Kanagasundaram et al., 2012). So, we have done modifications on weighted between-class scatter estimation. The weighted between-class scatter matrix, $\mathbf{S}_b^w$, is defined as

$$\mathbf{S}_b^w = \frac{1}{N}\sum_{i=1}^{S-1}\sum_{j=i+1}^{S} w(d_{ij}) n_i n_j (\overline{\mathbf{w}}_i - \overline{\mathbf{w}}_j)(\overline{\mathbf{w}}_i - \overline{\mathbf{w}}_j)^T, \tag{20}$$

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

In Eq. (20), the weighting function $w(d_{ij})$ is defined such that the classes that are closer to each other will be more heavily weighted. As we show in Appendix A, when $w(d_{ij})$ is set to 1, the weighted between-class scatter estimations will converge to the standard non-weighted between-class scatter from Eq. (8).

In this paper we are introducing the Euclidean distance, Mahalanobis distance and Bayes error weighting functions for speaker verification for the purpose of increasing the discriminant ability.

The Euclidean distance weighting function, $w_{(d_{ij})_{Euc}}$, can be defined as follows:

$$w_{(d_{ij})_{Euc}} = ((\overline{\mathbf{w}}_i - \overline{\mathbf{w}}_j)^T(\overline{\mathbf{w}}_i - \overline{\mathbf{w}}_j))^{-n}, \tag{21}$$

where $\overline{\mathbf{w}}_i$ and $\overline{\mathbf{w}}_j$ are the mean i-vectors of speaker i and j respectively, and n is a factor introduced to increase the separation for the classes that are closer. Classification performance will be analyzed with several arbitrary values of n. The Euclidean distance based weighting function is a monotonically decreasing function, so the classes that are closer together will be heavily weighted and classes that are away (outlier classes) will be lightly weighted to increase the discriminant ability.

The Mahalanobis distance, Δ_{ij}, between the means of classes i and j can be defined as,

$$\Delta_{ij} = \sqrt{(\overline{\mathbf{w}}_i - \overline{\mathbf{w}}_j)^T(\mathbf{S}_w)^{-1}(\overline{\mathbf{w}}_i - \overline{\mathbf{w}}_j)}, \tag{22}$$

where the within-class scatter matrix, $\mathbf{S}_w$, is estimated from Eq. (9). If the session i-vectors ($\mathbf{w}$) are uncorrelated in each speaker and are scaled to have unit variance, then $\mathbf{S}_w$ would be the identity matrix and the Mahalanobis distance will converge as the Euclidean distance between $\overline{\mathbf{w}}_i$ and $\overline{\mathbf{w}}_j$. We believe that there is some correlation between session i-vectors in each speaker and the within-class scatter is not an identity matrix. It can be shown that the presence of within-class scatter ($\mathbf{S}_w$) of $\mathbf{w}$ in the quadratic form in Eq. (22) allows for the different scales on which the variables are measured and for non-zero correlations between the variables.

The Mahalanobis distance weighting function, $w_{(d_{ij})_{Maha}}$, can be defined as follows:

$$w_{(d_{ij})_{Maha}} = (\Delta_{ij})^{-2n}. \tag{23}$$

where the Mahalanobis distance, Δ_{ij}, is estimated from Eq. (22). We introduce the Mahalanobis distance weighting function to i-vector speaker verification. It is also a monotonically decreasing function, so it will do the same job as Euclidean distance weighting function. In addition, it can be used to alleviate the dominant role of the outlier classes, so the Mahalanobis distance weighted between-class scatter has more discriminant ability than the Euclidean distance weighting function based weighted between-class scatter.

The final weighting parameter is based upon the Bayes error approximations of the mean accuracy amongst class pairs. The Bayes error based weighting function $w_{(d_{ij})_{Bayes}}$, can be calculated as:

$$w_{(d_{ij})_{Bayes}} = \frac{1}{2(\Delta_{ij})^2}\mathrm{Erf}\left(\frac{\Delta_{ij}}{2\sqrt{2}}\right), \tag{24}$$

where the Mahalanobis distance, Δ_{ij}, is estimated from Eq. (22). The Bayes error based weighting function is also used to heavily weight the classes that are very closer.

Once the weighted between-class scatter, $\mathbf{S}_b^w$, is estimated for the chosen weighting function, the standard within-class scatter $\mathbf{S}_w$ and the corresponding WLDA matrix ($\mathbf{A}$) can be estimated and applied as in traditional LDA. Finally, the WLDA channel compensated i-vector will be calculated using Eq. (12).

We also introduce the SN-WLDA approach to i-vector system, as an extension of the more basic SN-LDA approach, and analyze several source-dependent and source-independent weighting functions for i-vector speaker verification, which should show an improvement in performance across both matched and mismatched enrolment/verification conditions. Similarly to the SN-LDA between-class scatter calculations, the source normalized weighted between-class scatter matrix, $\mathbf{S}_b^{w_{src}}$, can be calculated as follows:

$$\mathbf{S}_b^{w_{src}} = \mathbf{S}_b^{w_{tel}} + \mathbf{S}_b^{w_{mic}}, \tag{25}$$

where the telephone-sourced dependent-weighted between-class scatter, $\mathbf{S}_b^{w_{tel}}$, and the microphone-sourced dependent-weighted between-class scatter, $\mathbf{S}_b^{w_{mic}}$, are individually calculated for telephone and microphone sources using Eq. (20).

We will be investigating the source-independent Euclidean distance weighting function (Eq. (21)), as it does not depend on any source variations. However, we will be investigating the source-dependent Mahalanobis distance and Bayes error weighting functions instead of source-independent weighting function, calculated using source-dependent within-class scatter variance to capture the source variation. The telephone and microphone source-dependent Mahalanobis distance, ${\Delta_{ij}}^{tel}$ and ${\Delta_{ij}}^{mic}$, can be defined as follows:

$${\Delta_{ij}}^{tel} = \sqrt{(\overline{\mathbf{w}}_i - \overline{\mathbf{w}}_j)^T (\mathbf{S}_w^{tel})^{-1} (\overline{\mathbf{w}}_i - \overline{\mathbf{w}}_j)}, \tag{26}$$

$${\Delta_{ij}}^{mic} = \sqrt{(\overline{\mathbf{w}}_i - \overline{\mathbf{w}}_j)^T (\mathbf{S}_w^{mic})^{-1} (\overline{\mathbf{w}}_i - \overline{\mathbf{w}}_j)}, \tag{27}$$

where $\mathbf{S}_w^{tel}$ and $\mathbf{S}_w^{mic}$ are telephone and microphone source-dependent within-class scatter matrices, individually calculated from telephone and microphone sources using Eq. (9). Once the source-dependent Mahalanobis distances, ${\Delta_{ij}}^{tel}$ and ${\Delta_{ij}}^{mic}$, are estimated from Eqs. (27) and (27), the source-dependent Mahalanobis distance and Bayes error weighting functions will be individually estimated from telephone and microphone sources using Eqs. (23) and (24).

In the SN-LDA algorithm, the within-class scatter matrix was estimated as the difference between total variance and the source-normalized between-class variance, but this approach is not taken for SN-WLDA, as the weighting parameters destroy the relationship between the total variance and the between-class scatter variance. For this reason, the within-class variance is estimated independently using Eq. (9) as in the LDA approach.

3.6. Real data scatter plot analysis

In this section, we will graphically observe how the original i-vector space and channel-compensated i-vector spaces separate the speakers. An overview of all seven channel compensation techniques alongside the raw i-vectors is shown in Fig. 1. All seven channel compensation techniques have been trained on whole development dataset, and the details of development set for channel compensation training is given in Section 4. We then randomly chose four representative speakers to project the original i-vector space into channel compensated reduced space using the channel compensation matrix. In channel compensation matrix estimation, the eigen-vectors were sorted in descending order according to corresponding eigen-values in order to illustrate the larger variation in Fig. 1.

It can be observed with the aid of Fig. 1(b) that WCCN projections scale a subspace in order to attenuate the high within-class variance. When we compare the WCCN and LDA projections with the aid of Fig. 1(b) and (c), it can be observed that LDA projection maximizes the between-speaker variability while minimizing the within speaker variability. After that when we observe the LDA and WLDA projections with the aid of Fig. 1(c) and (g), it can be clearly seen that WLDA projection increases the between speaker separability compared to LDA projections. Similarly to LDA and WLDA comparison, when we observe the SN-LDA and SN-WLDA projections with the aid of Fig. 1(d) and (h), it can be clearly seen that SN-WLDA projection increases the between speaker separability compared to SN-LDA projections.

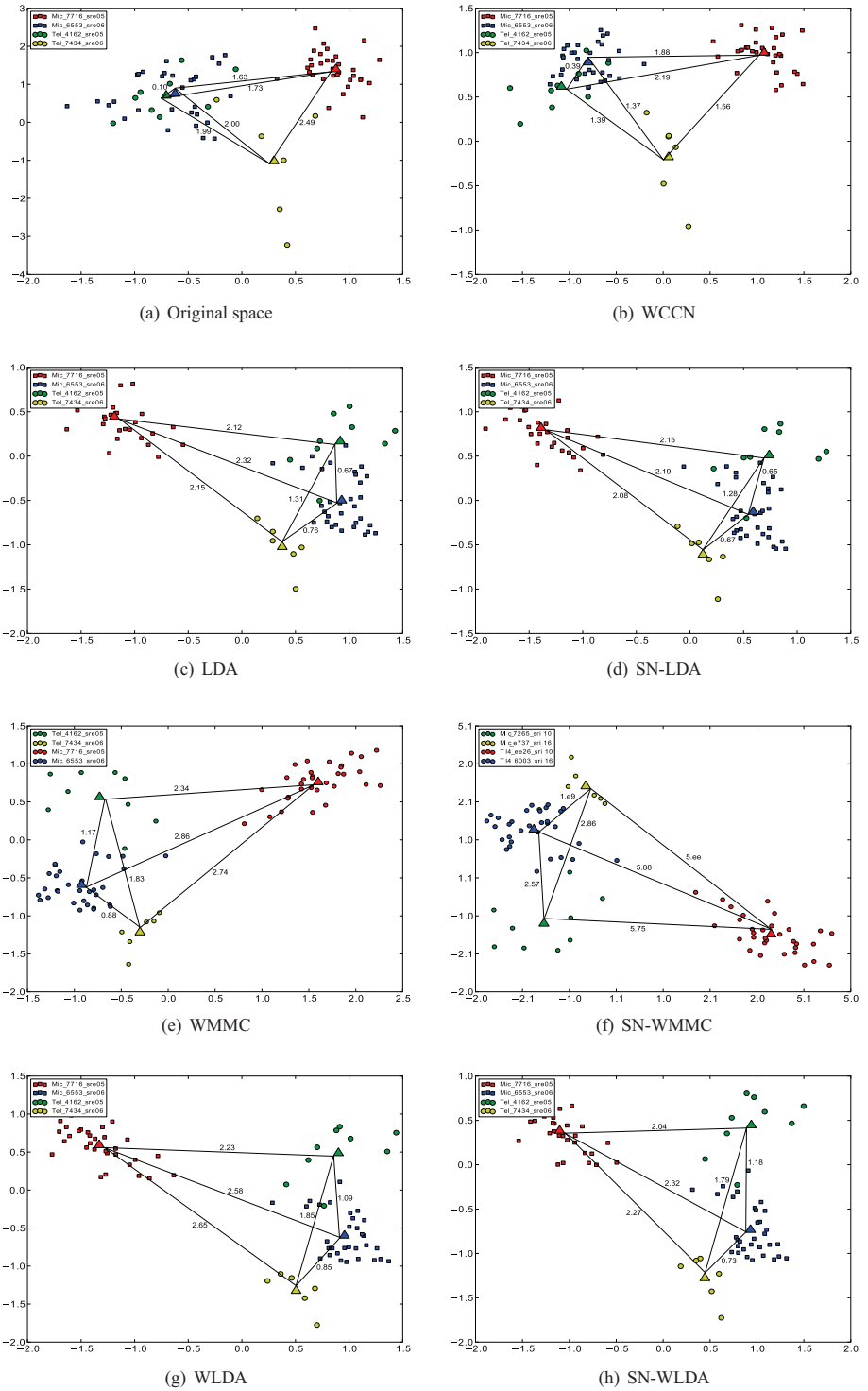


Fig. 1. Distribution of first two dimensions of female i-vectors features into (a) original space, or space projected using (b) WCCN, (c) LDA, (d) SN-LDA, (e) WMMC ($W = 0.25$), (f) SN-WMMC ($W = 0.25$), (g) WLDA ($Euc(n = 3)$) and (h) SN-WLDA ($Euc(n = 3)$).

3.7. Sequential channel compensation

In previous sections, we have detailed several individual channel compensation techniques. Individual LDA techniques are generally used to increase the inter-speaker variability while minimizing the intra-speaker variability, and WCCN approach is used to reduce the channel effect by minimizing the intra-speaker variability. Dehak et al. have found that the sequential approach of LDA followed by WCCN extracts more speaker discriminant features than individual LDA and WCCN approaches (Dehak et al., 2010), but continued research has found that any type of LDA followed by WCCN is generally considered the best approach (McLaren and van Leeuwen, 2011; Kanagasundaram et al., 2012). In the first stage of the WCCN[LDA] approach, LDA attempts to find a reduced set of axes $\mathbf{A}$ that minimizes the within-class variability while maximizing the between-class variability. The estimation of LDA ($\mathbf{A}$) was briefly described in Section 3.2.

In the second stage, WCCN is used as a channel compensation technique to scale a subspace in order to attenuate dimensions of high within-class variance. The WCCN transformation matrix ($\mathbf{B}_2$) is trained using the LDA-projected i-vectors from the first stage. The WCCN matrix ($\mathbf{B}_2$) is calculated using Cholesky decomposition of $\mathbf{B}_2\mathbf{B}_2{}^T = \mathbf{W}^{-1}$, where the within-class covariance matrix $\mathbf{W}$ is calculated using

$$\mathbf{W} = \frac{1}{S}\sum_{s=1}^{S}\sum_{i=1}^{n_s}(\mathbf{A}^T(\mathbf{w}_i^s - \overline{\mathbf{w}}_s))(\mathbf{A}^T(\mathbf{w}_i^s - \overline{\mathbf{w}}_s))^T \tag{28}$$

where $\mathbf{w}_i^s$ is the i-vector representation of i session of speaker s, the mean i-vector for each speaker ($\overline{\mathbf{w}}_s$) is equal to $(1/n_s)\sum_{i=1}^{n_s}\mathbf{w}_i^s$, S is the total number of speakers and n_s is number of utterances of speaker s.

The WCCN[LDA]-channel-compensated i-vector can be calculated as follows:

$$\hat{\mathbf{w}}_{LDA\rightarrow WCCN} = \mathbf{B}_2^T\mathbf{A}^T\mathbf{w} \tag{29}$$

The WCCN[LDA] approach is commonly used to compensate the channel variability in i-vector based speaker verification systems (Dehak et al., 2010). Similarly to the WCCN[LDA] approach outlined previously, we will also be investigating other channel compensation techniques, including SN-LDA, WMMC, SN-WMMC, WLDA and SN-WLDA followed by WCCN.

4. Experimental methodology

The i-vector based experiments were evaluated using the NIST 2008 and NIST 2010 Speaker Recognition Evaluation (SRE) corpora. Particulary, the NIST 2008 was used for parameter tuning task, and the NIST 2010 was used to validate the tuned parameters. For NIST 2008, the performance was evaluated using the equal error rate (EER) and the minimum decision cost function (DCF), calculated using $C_{miss} = 10$, $C_{FA} = 1$, and $P_{target} = 0.01$. NIST 2008 evaluation was performed using the *telephone–telephone*, *interview–interview*, *telephone–microphone* and *interview–telephone* enrolment-verification conditions (NIST, 2008). The performance for the NIST 2010 SRE was evaluated using the EER and the old minimum decision cost function (DCF_{old}), calculated using $C_{miss} = 10$, $C_{FA} = 1$, and $P_{target} = 0.01$, where evaluation was performed using the *telephone–telephone*, *interview–interview*, *interview–microphone* and *interview–telephone* condition (NIST, 2010).

We have used 13-dimensioned feature-warped MFCCs with appended delta coefficients and two gender-dependent universal background models (UBM) containing 512 Gaussian mixtures throughout our experiments. We kept the MFCC features dimension and number of UBM components in low values in order to reduce the computational cost, and it's easy to adapt to real world applications. UBMs were trained on telephone and microphone from NIST 2004, 2005, and 2006 SRE corpora for telephone and microphone i-vector experiments. These gender-dependent UBMs were used to calculate the Baum-Welch statistics before training a gender dependent total-variability subspace of dimension $R_w = 500$, which was then used to calculate the i-vector speaker representations. Total variability representation, channel compensation matrices and length-normalized GPLDA model 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 (dimension of U_1) equal to 120 as best value according to speaker verification performance. A full precision matrix was used for $\mathbf{\Lambda}$, rather than the diagonal. ZT normalization was applied to telephone and microphone speech based CSS i-vector system experiments and S normalization was applied

Table 1
Comparison of i-vector approach performance with/without standard channel compensation techniques on the common set of the 2008 NIST SRE short2–short3 conditions. The best performing systems by both EER and DCF are highlighted across each row.

System	*Interview–interview*		*Interview–telephone*		*Telephone–microphone*		*Telephone–telephone*	
	EER (%)	DCF	EER (%)	DCF	EER (%)	DCF	EER (%)	DCF
Individual approach								
Raw i-vectors	11.09	0.0522	14.10	0.0505	9.44	0.0362	5.68	0.0255
WCCN	6.84	0.0357	7.74	0.0356	5.70	0.0239	3.71	0.0166
LDA	6.94	0.0328	8.03	0.0379	7.06	0.0283	3.95	0.0178
SN-LDA	7.20	0.0330	7.83	0.0382	6.93	0.0286	3.87	0.0170
Sequential approach								
WCCN[LDA]	**4.61**	**0.0228**	5.99	0.0293	5.10	0.0222	**2.80**	**0.0134**
WCCN[SN-LDA]	4.73	0.0235	**5.90**	**0.0278**	**4.83**	**0.0208**	2.96	0.0136

to length-normalized GPLDA system experiments. Randomly selected telephone and microphone utterances from NIST 2004, 2005 and 2006 were pooled to form the ZT and S normalization dataset. For the NIST 2008 evaluation, in most of the cases, the system achieved the best performance, when the channel compensation approach dimension was selected as 150. For NIST 2010 evaluation, we have also chosen the channel compensation approach dimension as 150, in order to show that the best value for NIST 2008 evaluation is robust to other dataset as well.

Score-level fusion is implemented using the FoCal toolkit (Brummer, 2005) to optimize linear regression parameters. The fusion weights were learned using scores from the NIST 2008 short2–short3 conditions.

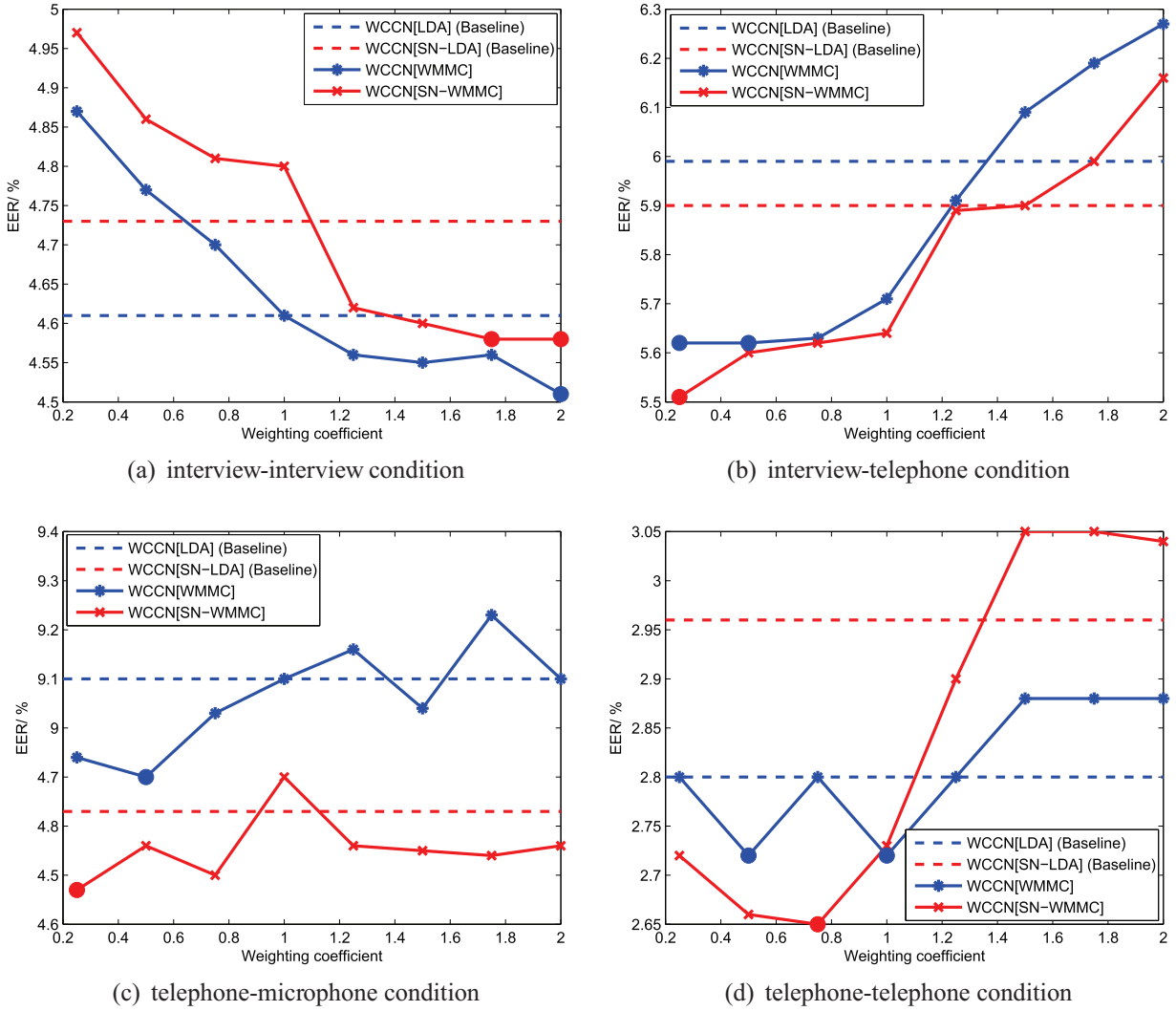
5. Results and discussion

In this paper, initially, we define the channel compensation approaches, including WCCN, LDA and SN-LDA as unweighted channel compensation approaches, as they do not depend on any weighting coefficients. However, we define the channel compensation approaches, including WLDA, SN-WLDA, WMMC, SN-WMMC as weighted channel compensation approaches, as they depend on weighting coefficients. Initial experiments were conducted without channel compensation techniques (raw i-vectors) and with unweighted channel compensation techniques, including WCCN, LDA and SN-LDA. Unweighted channel compensation techniques were analyzed both with and without WCCN. Following this, several weighted channel compensation techniques will be analyzed in combination with WCCN to identify the best overall channel compensation approach. After that several channel compensation techniques were analyzed to combine through score-level fusion to illustrate the complementary nature of the channel compensation techniques. Finally, the best channel compensation approach was investigated with length-normalized GPLDA system.

5.1. Unweighted channel compensation techniques

Speaker verification experiments were conducted with individual channel compensation techniques, and in combination with WCCN (as motivated by Dehak et al. (2010)) to see how channel compensated i-vectors perform over raw uncompensated i-vectors. Table 1 presents the results from these experiments on the common set of the 2008 NIST SRE short2–short3 conditions. The results show that channel compensation can achieve major improvement over the raw i-vector approach. If we have a closer look at the individual channel compensation techniques, it can be clearly seen that WCCN performs better than LDA and SN-LDA as channel variations mainly depend on the within-speaker variation than between-speaker variation.

Further, if we look at the channel compensation techniques in combination with WCCN, we find improved performance over individual channel compensation systems, which supports the findings of Dehak et al. (2010). Based upon the results shown here, and similar findings by Dehak et al. (2010) and McLaren and van Leeuwen (2011), it is clear that best performance can be obtained by accompanying more sophisticated channel compensation techniques with WCCN, and this is the approach that will be taken throughout the reminder of experiments in this paper.



(a) interview-interview condition

(b) interview-telephone condition

(c) telephone-microphone condition

(d) telephone-telephone condition

Fig. 2. Comparison of EER values of WCCN[WMMC] and WCCN[SN-WMMC] approaches at different weighting coefficients in different enrolment and verification conditions.

5.2. Training weighted channel compensation techniques

Before the weighted channel compensated techniques WMMC and WLDA (as well as SN-WMMC and SN-WLDA) can be evaluated against the traditional LDA (and SN-LDA) approaches, the best configuration of these techniques must be determined. The NIST 2008 data set was used to find the best configuration.

5.2.1. Choosing the WMMC weighting coefficient

The WMMC and SN-WMMC approaches have the flexibility to change the importance of the within- and between-class scatters, and those performance were analyzed at different levels of the influence of within-class scatter ($\mathbf{S}_w$) based on manual weighting coefficients (W) in Eq. (18). The EER performance of WCCN[WMMC] and WCCN[SN-WMMC] across different train-test sources at different weighting coefficients is shown in Fig. 2. It can be clearly seen with the aid of Fig. 2(b) and (d) that when the weighting coefficient is increased around above 1, and therefore the level of influence of within-class scatter is increased, telephone speech verification condition performance goes down below baseline performance, suggesting that, for this condition, the within- and between-class scatter variances are equally important. However, when the level of influence of within-class scatter is increased around above 1, the system

achieves better performance than baseline on *interview–interview* condition (Fig. 2(a)), as the within-class scatter variance plays a major role in the higher channel variation present in interview speech. The best values of WMMC weighting coefficients for all conditions were highlighted using a larger circle symbol in Fig. 2, and these values will be used in future experiments within this paper.

5.2.2. Choosing the WLDA weighting functions

The importance of weighted between-class scatters on LDA and SN-LDA estimations were analyzed in this section. The performance of these approaches were analyzed with respect to these weighting functions: Bayes error, Euclidean distance and Mahalanobis distance. While Bayes error weighting function is not a parameterized approach, the Euclidean and Mahalanobis distance functions are constructed as monotonically decreasing functions, where the n is used to change the sensitivity of the weighting function to the underlying distance, where a higher n value indicates more sensitivity. The Euclidean and Mahalanobis distance weighting functions were analyzed at different n values to see the effect on between-speaker separability. This analysis is shown in Fig. 3 for WLDA and Fig. 4 for SN-WLDA.

It can be clearly seen with the aid of both Figs. 3 and 4 that when the n value increases above some level, around 4 for WLDA and 2 for SN-WLDA, the performance goes down in all enrolment and verification conditions, as the weighting functions with higher n value reduces the quality of between-class scatter variance. The weighting functions with higher n values fail to alleviate the dominant role of the outlier classes. If we closely look at on interview and microphone speech verification conditions (Figs. 3(a), (c) and 4(a), (c)), the WLDA and SN-WLDA approaches achieved better performance than baseline systems over the wide range of n value's choice. Even though the Bayes error weighting function is a non-parametric approach, the Bayes error based WLDA and SN-WLDA approaches achieved reasonably better performance over baseline approaches.

5.3. Comparing all techniques

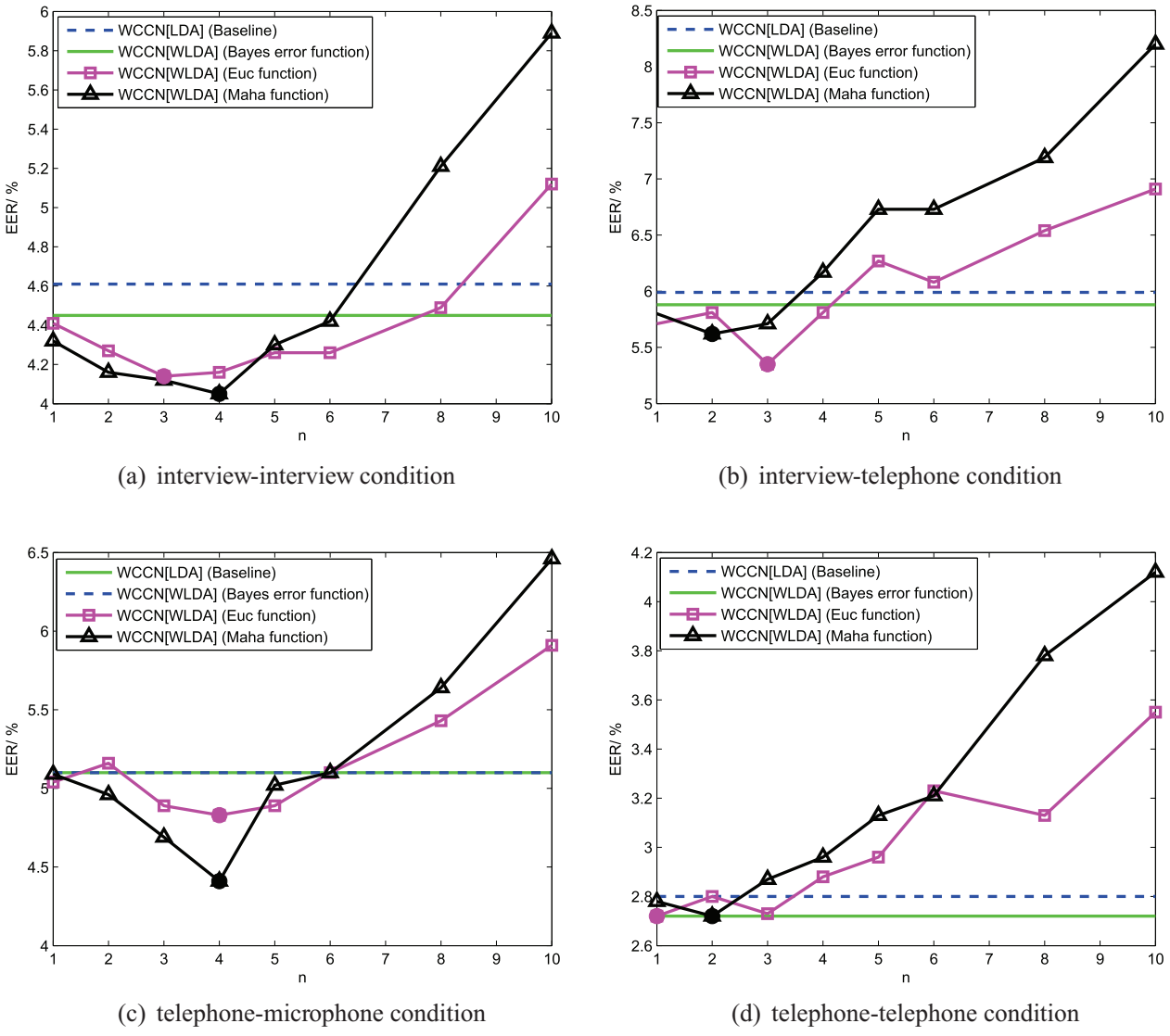
Weighted channel compensation techniques were finely tuned in the previous section. In this section, weighted and unweighted channel compensation techniques are compared to identify the best channel compensation approach. Table 2(a) and (b) presents the results comparing the performance of WCCN[WMMC] and WCCN[WLDA] against the baseline system, WCCN[LDA], on the common set of the 2008 NIST SRE short2–short3 and 2010 NIST SRE core–core conditions. The WCCN[WMMC] and WCCN[WLDA] results were presented with optimized weighting parameters, as detailed in the previous section.

Initially, if we compare the performance between the WMMC and LDA approaches on NIST 2008 short2–short3 condition, the WMMC technique achieved over 2% relative improvement in EER over LDA on all training and testing conditions, by finely tuning the required influence of within- and between-class scatter variances. However, the WMMC technique hasn't shown consistent improvement over LDA on NIST 2010 core–core condition, as the required influence of within- and between-class scatter variances were finely selected from NIST 2008 data set.

Secondly, it can be clearly seen that, by taking advantage of the speaker discriminative information, the WLDA techniques have shown over 8% improvement in EER on NIST 2008 interview and microphone speech verification conditions compared to the LDA approach. The WLDA techniques have also shown 10% improvement in EER on NIST 2008 *interview–telephone* condition over the LDA approach. The WLDA techniques have not shown great improvement over LDA and WMMC in *telephone–telephone* condition, because most of the telephone-speech speaker means are closely situated and equally distributed due to channel similarities. When we compare the performance of WLDA approaches against baseline, LDA approach, on NIST 2010 core–core condition, there is an improvement, but further improvements can be achieved, if the weighting functions coefficients and LDA dimension were selected from NIST 2010 dataset.

In Table 3(a) and (b), we take advantage of source-normalization (SN), and present the results comparing the performance of WCCN[SN-WMMC] and WCCN[SN-WLDA] against the baseline system, WCCN[SN-LDA], on the common set of the 2008 NIST SRE short2–short3 and 2010 NIST SRE core–core conditions. The WCCN[SN-WMMC] and WCCN[SN-WLDA] results were presented with optimized weighting parameters, as detailed in the previous section.

Similarly to Table 2, it can be clearly seen that, by capturing the source variation as well as finely tuning the influence of within- and between-class scatter variations, the SN-WMMC technique does show over 3% improvement in EER for NIST 2008 interview and microphone verification and over 6% improvement in EER for NIST 2008 telephone



(a) interview-interview condition

(b) interview-telephone condition

(c) telephone-microphone condition

(d) telephone-telephone condition

Fig. 3. Comparison of EER values of WCCN[WLDA] approach based on Euclidean and Mahalanobis distance weighting functions at different n values in different enrolment and verification conditions. Note that in (c), the baseline and Bayes error curves overlap and cannot be visually separated.

verification over the SN-LDA approach. However, the SN-WMMC technique has not shown consistent improvement over SN-LDA on NIST 2010 core–core condition, as the required influence of within- and between-class scatter variances were finely selected from NIST 2008 data set.

When we compare the performance of SN-WLDA to SN-LDA, it can be clearly seen that, by extracting the discriminatory information between pairs of speakers as well as capturing the source variation information, the Mahalanobis distance-based SN-WLDA shows over 20% improvement in EER for NIST 2008 interview and microphone verification and over 10% improvement in EER for NIST 2008 telephone speech verification. If we closely look at the SN-WLDA approach with several weighting functions, the Mahalanobis distance-based SN-WLDA showed greater improvement over the Euclidean distance-based SN-WLDA, as the Mahalanobis distance weighting function was used to alleviate the dominant role of the outlier classes as well as it was calculated based on source dependent within-class scatter variance and it has more speaker discriminant information. The Bayes error weighting function is also based on source-dependent within-class scatter variance, however, it hasn't shown improvement over Mahalanobis distance-based SN-WLDA as it is a non-parametric weighting function. If we compare the SN-WLDA approach

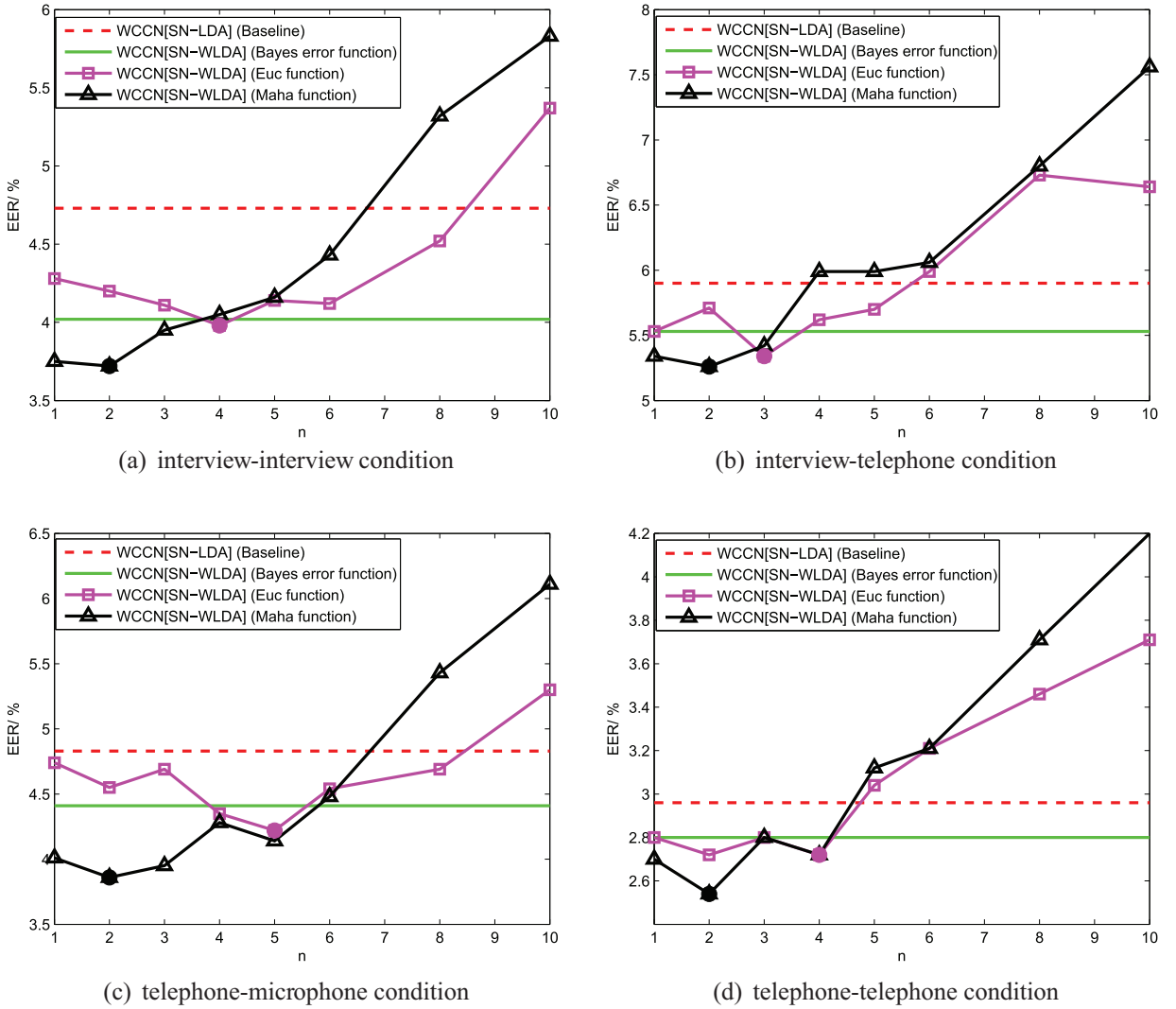


Fig. 4. Comparison of EER values of WCCN[SN-WLDA] approach based on Euclidean, Mahalanobis distance weighting functions at different n values in different enrolment and verification conditions.

against baseline approach, SN-LDA, the SN-WLDA approach shows over 10% improvement in EER on NIST 2010 *interview–interview* and *interview–microphone* conditions. The improvements over baseline suggests that the optimal parameter values are robust for other dataset as well. However, if we select the optimal parameters on same data set by looking the performance, the performance would be better than when optimal parameters are trained on different data set.

In overall, when the performance of WLDA technique is compared with SN-WLDA technique (refer to Tables 2 and 3), the SN-WLDA achieved better performance than the WLDA in all the enrolment and verification conditions, as the SN-WLDA approach captures the source variation information and also extracts the discriminatory information between pairs of classes.

5.4. Score-level fusion channel compensation analysis

Several novel channel compensation techniques, including WMMC, SN-WMMC, WLDA and SN-WLDA were investigated in combination with WCCN previously. However, multiple channel compensation approaches combined using score-level fusion to extract speaker complementary information has not yet been investigated. In this section,

Table 2
Comparison of WCCN[WMMC] and WCCN[WLDA] systems against the WCCN[LDA] system on the common set of the 2008 NIST SRE short2–short3 and 2010 NIST SRE core–core conditions. The best performing systems by both EER and DCF are highlighted down each column.

System	*Interview–interview*		*Interview–telephone*		*Telephone–microphone*		*Telephone–telephone*	
	EER (%)	DCF	EER (%)	DCF	EER (%)	DCF	EER (%)	DCF
(a) NIST 2008 short2–short3 condition								
Baseline system								
WCCN[LDA]	4.61	0.0228	5.99	0.0293	5.10	0.0222	2.80	0.0134
Weighted MMMC system								
WCCN[WMMC]	4.51	0.0231	5.62	**0.0287**	4.90	0.0223	**2.72**	0.0135
Weighted LDA system								
WCCN[WLDA(Bayes)]	4.45	0.0221	5.88	0.0295	5.10	0.0221	**2.72**	0.0132
WCCN[WLDA(Euc)]	4.14	0.0199	**5.35**	**0.0287**	4.89	**0.0213**	2.73	**0.0128**
WCCN[WLDA(Maha)]	**4.05**	**0.0198**	5.62	0.0291	**4.69**	0.0218	**2.72**	0.0130

System	*Interview–interview*		*Interview–telephone*		*Interview–microphone*		*Telephone–telephone*	
	EER (%)	DCF_{old}	EER (%)	DCF_{old}	EER (%)	DCF_{old}	EER (%)	DCF_{old}
(b) NIST 2010 core–core condition								
Baseline system								
WCCN[LDA]	7.13	0.0295	5.45	0.0240	4.27	0.0198	3.81	0.0154
Weighted MMMC system								
WCCN[WMMC]	7.25	0.0311	5.45	0.0256	4.24	0.0199	**3.54**	0.0173
Weighted LDA system								
WCCN[WLDA(Bayes)]	7.10	0.0292	5.39	0.0239	4.22	**0.0197**	3.81	0.0153
WCCN[WLDA(Euc)]	6.97	**0.0290**	5.33	**0.0238**	4.27	0.0201	3.83	**0.0152**
WCCN[WLDA(Maha)]	**6.85**	0.0291	**5.27**	0.0239	**3.97**	0.0201	4.10	0.0153

the score-level fused approach is investigated to combine all the source-normalize channel compensation approaches, including SN-LDA, SN-WMMC and SN-WLDA to extract the complementary speaker information.

Fusion system has shown improvement in all conditions except *telephone–telephone* condition. If we look at Tables 2 and 3, it is clear that each individual system has not shown much improvement on *telephone–telephone* condition. So, it is unlikely to expect improvement on fusion results on *telephone–telephone* condition. We have chosen the *interview–telephone* condition to analyze the score-level fusion approach, as *interview–telephone* condition has shown least performance over other enrolment and verification conditions in the previous experiments. Table 4 presents results comparing the performance of score-level fused approaches on common set of NIST 2008 short2–short3 and NIST 2010 core–core *interview-telephone* conditions. We have used NIST 2008 short2–short3 condition scores to tune the fusion weights. The score fused system has shown improvement over individual systems on both NIST 2008 short2–short3 and NIST 2010 core–core *interview–telephone* conditions, which suggests that the fused weights are not optimistically biased for a given corpus. For score fusion experiments, initially we have fused all the source-normalized channel compensation approaches, and each step we cut off the least contribution system. By doing this approach, we have found WCCN[SN-WMMC] and WCCN[SN-WLDA(Maha)] as the two best systems to fuse together. For NIST 2010 evaluations, the weighted channel compensation approaches, including SN-WMMC and SN-WLDA were trained using same optimized parameters, which were obtained from Figs. 2 and 4. The improvements over baseline suggest that the optimal parameter values are robust for other dataset as well.

It is also clear that all the source-normalized channel compensation approaches fused system has shown over 8% improvement in DCF over the best single approach, WCCN[SN-WLDA(Maha)], on NIST 2008 short2–short3 *interview–telephone* condition, as all the source-normalized fused system extracts complementary speaker information. If we closely look at the fusion weights, the contribution of WCCN[SN-WMMC] approach is greater compared to weighting functions based WCCN[SN-WLDA], as all the weighting functions based WCCN[SN-WLDA] approaches are correlated, and the WCCN[SN-WMMC] approach has more complementary speaker information.

Table 3
Comparison of WCCN[SN-WMMC] and WCCN[SN-WLDA] systems against the WCCN[SN-LDA] system on the common set of the 2008 NIST SRE short2–short3 and 2010 NIST SRE core–core conditions. The best performing systems by both EER and DCF are highlighted down each column.

System	*Interview–interview*		*Interview–telephone*		*Telephone–microphone*		*Telephone–telephone*	
	EER (%)	DCF	EER (%)	DCF	EER (%)	DCF	EER (%)	DCF
(a) NIST 2008 short2–short3 condition								
Baseline system								
WCCN[SN-LDA]	4.73	0.0235	5.90	0.0278	4.83	0.0208	2.96	0.0136
Source-normalized WMMC system								
WCCN[SN-WMMC]	4.58	0.0231	5.51	0.0266	4.67	0.0206	2.65	0.0136
Source-normalized WLDA system								
WCCN[SN-WLDA(Bayes)]	4.02	0.0196	5.53	0.0251	4.41	0.0184	2.80	0.0130
WCCN[SN-WLDA(Euc)]	3.98	0.0190	5.34	0.0262	4.22	0.0203	2.72	0.0130
WCCN[SN-WLDA(Maha)]	**3.72**	**0.0178**	**5.26**	**0.0249**	**3.86**	**0.0179**	**2.54**	**0.0125**

System	*Interview–interview*		*Interview–telephone*		*Interview–microphone*		*Telephone–telephone*	
	EER (%)	DCF_{old}	EER (%)	DCF_{old}	EER (%)	DCF_{old}	EER (%)	DCF_{old}
(b) NIST 2010 core–core condition								
Baseline system								
WCCN[SN-LDA]	7.27	0.0302	5.02	0.0239	4.52	0.0202	3.78	0.0155
Source-normalized WMMC system								
WCCN[SN-WMMC]	7.29	0.0294	5.20	0.0238	4.56	0.0203	3.95	**0.0154**
Source-normalized WLDA system								
WCCN[SN-WLDA(Bayes)]	6.61	0.0280	**4.59**	0.0217	4.02	**0.0193**	3.68	0.0155
WCCN[SN-WLDA(Euc)]	6.85	0.0288	4.72	0.0225	**3.85**	0.0198	3.94	0.0165
WCCN[SN-WLDA(Maha)]	**6.44**	**0.0272**	4.66	**0.0210**	3.98	0.0194	**3.67**	0.0156

Table 4
Comparison of score-level fusion systems on the common set of the NIST 2008 SRE short2–short3 and NIST 2010 SRE core–core *interview–telephone* conditions. The best performing systems by both EER and DCF are highlighted down each column.

Fused system ($a_1S_1 + a_2S_2 + a_3S_3 + a_4S_4 + a_5S_5 + b$)	**Focal weights tuned on 2008**								
WCCN[SN-LDA] (a_1)	1.00	–	–	–	–	−0.66	–	–	–
WCCN[SN-WMMC] (a_2)	–	1.00	–	–	–	1.38	0.86	0.98	1.21
WCCN[SN-WLDA(Bayes)] (a_3)	–	–	1.00	–	–	0.46	0.33	–	–
WCCN[SN-WLDA(Euc)] (a_4)	–	–	–	1.00	–	0.54	0.49	0.52	–
WCCN[SN-WLDA(Maha)] (a_5)	–	–	–	–	1.00	1.07	1.12	1.30	1.57
Constant (b)	–	–	–	–	–	−5.36	−5.37	−5.35	−5.29
NIST 2008 SRE short2–short3 *interview–telephone* condition									
EER (%)	5.90	5.51	5.53	5.34	5.26	**5.16**	5.26	5.26	5.34
DCF	0.0278	0.0266	0.0251	0.0262	0.0249	**0.0230**	0.0235	0.0237	0.0235
NIST 2010 SRE core–core *interview–telephone* condition									
EER (%)	5.02	5.20	4.59	4.72	4.66	4.59	**4.47**	4.48	**4.47**
DCF_{old}	0.0239	0.0238	0.0217	0.0225	0.0210	**0.0207**	0.0208	0.0208	0.0208

5.5. Length-normalized GPLDA analysis on channel compensated i-vector features

Several novel channel compensation approaches were analyzed with CSS based i-vector system in previous sections. We have also found that SN-WLDA approach is the best channel compensation approach when comparing with WMMC, WLDA and SN-WLDA approaches. In this section, we have analyzed that how the SN-WLDA projected length-normalized GPLDA system performs over the baseline approaches, LDA and SN-LDA projected length-normalized GPLDA systems. Table 5(a) and (b) presents the results on the common set of the NIST SRE 2008 short2–short3 and NIST SRE 2010 core–core conditions. If we compare the SN-WLDA projected GPLDA against baseline approach,

Table 5
Comparison of SN-WLDA projected length-normalized GPLDA system against the standard length-normalized GPLDA, WCCN[LDA] and WCCN[SN-LDA] projected length-normalized GPLDA systems on the common set of the 2008 NIST SRE short2–short3 and 2010 NIST SRE core–core conditions. The best performing systems by both EER and DCF are highlighted down each column.

System	*Interview–interview*		*Interview–telephone*		*Telephone–microphone*		*Telephone–telephone*	
	EER (%)	DCF	EER (%)	DCF	EER (%)	DCF	EER (%)	DCF
(a) NIST 2008 short2–short3 condition								
Baseline system								
Standard GPLDA	5.05	0.0264	5.43	0.0275	4.08	0.0204	2.63	0.0136
WCCN[LDA]-GPLDA	4.29	0.0214	5.51	0.0254	4.35	0.0195	2.63	0.0126
WCCN[SN-LDA]-GPLDA	4.15	0.0210	5.25	0.0249	3.88	0.0189	2.72	0.0124
SN-WLDA projected length-normalized GPLDA system								
WCCN[SN-WLDA(Bayes)]-GPLDA	3.91	0.0189	**4.96**	0.0233	3.81	0.0171	**2.39**	**0.0118**
WCCN[SN-WLDA(Euc)]-GPLDA	3.89	0.0196	5.27	**0.0227**	**3.73**	0.0174	2.47	0.0124
WCCN[SN-WLDA(Maha)]-GPLDA	**3.61**	**0.0174**	5.16	0.0228	3.74	**0.0157**	2.47	0.0119

System	*Interview–interview*		*Interview–telephone*		*Interview–microphone*		*Telephone–telephone*	
	EER (%)	DCF_{old}	EER (%)	DCF_{old}	EER (%)	DCF_{old}	EER (%)	DCF_{old}
(b) NIST 2010 core–core condition								
Baseline system								
Standard GPLDA	7.21	0.0338	4.84	0.0239	4.56	0.0244	3.39	0.0167
WCCN[LDA]-GPLDA	6.76	0.0292	4.41	0.0220	4.10	0.0196	3.41	0.0152
WCCN[SN-LDA]-GPLDA	6.91	0.0299	4.41	0.0212	4.15	0.0200	3.51	0.0152
SN-WLDA projected length-normalized GPLDA system								
WCCN[SN-WLDA(Bayes)]-GPLDA	6.27	0.0274	4.36	0.0205	3.76	0.0190	3.39	0.0152
WCCN[SN-WLDA(Euc)]-GPLDA	6.37	0.0285	4.35	0.0202	**3.38**	0.0190	3.56	0.0144
WCCN[SN-WLDA(Maha)]-GPLDA	**5.94**	**0.0262**	**4.10**	**0.0193**	3.43	**0.0182**	**3.25**	**0.0143**

SN-LDA projected GPLDA, SN-WLDA projected GPLDA system shows over 14% improvement in EER for NIST SRE 2010 interview and microphone verification and over 7% improvement in EER for NIST SRE 2010 telephone verification, as it extracts the discriminatory information between pairs of speakers as well as capturing the source variation information.

Based upon all the experiments on NIST 2008 and NIST 2010 evaluations, we believe that improvements demonstrated throughout this paper of advanced channel compensation techniques for CSS-based i-vector speaker representation can also translate well into the length-normalized GPLDA approach.

6. Conclusion

In this paper, we have analyzed advanced channel compensation techniques for the purpose of improving i-vector speaker verification performance in the presence of high intersession variability using the NIST 2008 and 2010 SRE corpora. Firstly, we have introduced the WMMC as an alternative to LDA, that can provide additional flexibility to change the relative influence of the within- and between-class variances. With the added benefit of source-normalization, the SN-WMMC technique has shown an improvement for all verification conditions. Secondly, we have introduced the WLDA technique, based upon the weighted pairwise Fisher criterion. Further, by extracting the discriminatory information between pairs of speakers as well as capturing the source variation information in the development i-vector space, the SN-WLDA has shown over 20% improvement in EER for NIST 2008 interview and microphone verification and over 10% improvement in EER for NIST 2008 telephone verification, when compared to SN-LDA. Further, score-level fusion techniques were analyzed to combine the best channel compensation approaches, to show over 8% improvement in DCF over the best single approach, (SN-WLDA), for NIST 2008 *interview–telephone* condition. Finally, the SN-WLDA projected length-normalized GPLDA system shows over 14% improvement in EER for NIST SRE 2010 interview and microphone verification and over 7% improvement in EER for NIST SRE 2010 telephone verification when compared to SN-LDA projected length-normalized GPLDA system, as it models the channel variation

in GPLDA space as well as it extracts the discriminatory information between pairs of speakers and captures the source variation information. In our future research, the proposed techniques will be modified to apply within PLDA model development.

Acknowledgements

This research was supported by an Australian Research Council (ARC) Discovery grant DP0877835. The authors also thank the reviewers for their valuable comments which has enabled us to significantly improve the quality of the paper.

Appendix A. Weighted between-class scatter estimation with unity weighting function

Weighted between-class scatter matrix can be calculated as follows:

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

When weighting function $w(d_{ij})$ equals to 1, weighted between-class scatter equation can be written as follows:

$$\mathbf{S}_b^w = \frac{1}{N}\sum_{i=1}^{S-1}\sum_{j=i+1}^{S} n_i n_j (\overline{\mathbf{w}}_i - \overline{\mathbf{w}}_j)(\overline{\mathbf{w}}_i - \overline{\mathbf{w}}_j)^T,$$

$$\begin{aligned}\mathbf{S}_b^w = \frac{1}{2N}(&2n_1n_2(\overline{\mathbf{w}}_1 - \overline{\mathbf{w}}_2)(\overline{\mathbf{w}}_1 - \overline{\mathbf{w}}_2)^T + 2n_1n_3(\overline{\mathbf{w}}_1 - \overline{\mathbf{w}}_3)(\overline{\mathbf{w}}_1 - \overline{\mathbf{w}}_3)^T + \cdots + 2n_1n_s(\overline{\mathbf{w}}_1 - \overline{\mathbf{w}}_s)(\overline{\mathbf{w}}_1 - \overline{\mathbf{w}}_s)^T \\ &+ 2n_2n_3(\overline{\mathbf{w}}_2 - \overline{\mathbf{w}}_3)(\overline{\mathbf{w}}_2 - \overline{\mathbf{w}}_3)^T + 2n_2n_4(\overline{\mathbf{w}}_2 - \overline{\mathbf{w}}_4)(\overline{\mathbf{w}}_2\overline{\mathbf{w}}_4)^T + \cdots + 2n_2n_s(\overline{\mathbf{w}}_2 - \overline{\mathbf{w}}_s)(\overline{\mathbf{w}}_2 - \overline{\mathbf{w}}_s)^T \\ &+ \cdots + 2n_{s-1}n_s(\overline{\mathbf{w}}_{s-1} - \overline{\mathbf{w}}_s)(\overline{\mathbf{w}}_{s-1} - \overline{\mathbf{w}}_s)^T)\end{aligned}$$

$$\begin{aligned}\mathbf{S}_b^w = \frac{1}{2N}(&n_1n_1(\overline{\mathbf{w}}_1 - \overline{\mathbf{w}}_1)(\overline{\mathbf{w}}_1 - \overline{\mathbf{w}}_1)^T + n_1n_2(\overline{\mathbf{w}}_1 - \overline{\mathbf{w}}_2)(\overline{\mathbf{w}}_1 - \overline{\mathbf{w}}_2)^T + \cdots + n_1n_s(\overline{\mathbf{w}}_1 - \overline{\mathbf{w}}_s)(\overline{\mathbf{w}}_1 - \overline{\mathbf{w}}_s)^T \\ &+ n_2n_1(\overline{\mathbf{w}}_2 - \overline{\mathbf{w}}_1)(\overline{\mathbf{w}}_2 - \overline{\mathbf{w}}_1)^T + n_2n_2(\overline{\mathbf{w}}_2 - \overline{\mathbf{w}}_2)(\overline{\mathbf{w}}_2 - \overline{\mathbf{w}}_2)^T + \cdots + n_2n_s(\overline{\mathbf{w}}_2 - \overline{\mathbf{w}}_s)(\overline{\mathbf{w}}_2 - \overline{\mathbf{w}}_s)^T + \cdots \\ &+ n_sn_1(\overline{\mathbf{w}}_s - \overline{\mathbf{w}}_1)(\overline{\mathbf{w}}_s - \overline{\mathbf{w}}_1)^T + n_sn_2(\overline{\mathbf{w}}_s - \overline{\mathbf{w}}_2)(\overline{\mathbf{w}}_s - \overline{\mathbf{w}}_2)^T + \cdots + n_sn_s(\overline{\mathbf{w}}_s - \overline{\mathbf{w}}_s)(\overline{\mathbf{w}}_s - \overline{\mathbf{w}}_s)^T)\end{aligned}$$

$$\mathbf{S}_b^w = \frac{1}{2N}\sum_{i=1}^{S}\sum_{j=i}^{S} n_i n_j (\overline{\mathbf{w}}_i - \overline{\mathbf{w}}_j)(\overline{\mathbf{w}}_i - \overline{\mathbf{w}}_j)^T$$

$$\mathbf{S}_b^w = \frac{1}{2N}\sum_{i=1}^{S}\sum_{j=1}^{S} n_i n_j ((\overline{\mathbf{w}}_i - \overline{\mathbf{w}}) + (\overline{\mathbf{w}} - \overline{\mathbf{w}}_j)) \times ((\overline{\mathbf{w}}_i - \overline{\mathbf{w}}) + (\overline{\mathbf{w}} - \overline{\mathbf{w}}_j))^T$$

$$\mathbf{S}_b^w = \frac{1}{2N}\sum_{i=1}^{S}\sum_{j=1}^{S} n_i n_j ((\overline{\mathbf{w}}_i - \overline{\mathbf{w}})(\overline{\mathbf{w}}_i - \overline{\mathbf{w}})^T + (\overline{\mathbf{w}}_i - \overline{\mathbf{w}})(\overline{\mathbf{w}} - \overline{\mathbf{w}}_j)^T + (\overline{\mathbf{w}} - \overline{\mathbf{w}}_j)(\overline{\mathbf{w}}_i - \overline{\mathbf{w}})^T + (\overline{\mathbf{w}} - \overline{\mathbf{w}}_j)(\overline{\mathbf{w}} - \overline{\mathbf{w}}_j)^T)$$

Since $\sum_{i=1}^{S}(n_i/N) = 1$, we can combine the first and last outer product terms above to get

$$\mathbf{S}_b^w = \sum_{i=1}^{S} n_i (\overline{\mathbf{w}}_i - \overline{\mathbf{w}})(\overline{\mathbf{w}}_i - \overline{\mathbf{w}})^T + \frac{1}{2N}\sum_{i=1}^{S}\sum_{j=1}^{S} n_i n_j (\overline{\mathbf{w}}_i - \overline{\mathbf{w}})(\overline{\mathbf{w}} - \overline{\mathbf{w}}_j)^T + \frac{1}{2N}\sum_{i=1}^{S}\sum_{j=1}^{S} n_i n_j (\overline{\mathbf{w}}_j - \overline{\mathbf{w}})(\overline{\mathbf{w}} - \overline{\mathbf{w}}_i)^T$$

Examine the last two terms above, we note that $\sum_{i=1}^{S}(n_i/N)\overline{\mathbf{w}}_i = \overline{\mathbf{w}}$ and therefore $\sum_{i=1}^{S}(n_i/N)(\overline{\mathbf{w}} - \overline{\mathbf{w}}_i) = 0$. Weighted between-class scatter will converge as follows:

$$\mathbf{S}_b^w = \sum_{i=1}^{S} n_i(\overline{\mathbf{w}}_i - \overline{\mathbf{w}})(\overline{\mathbf{w}}_i - \overline{\mathbf{w}})^T$$

References

Auckenthaler, R., Carey, M., Lloyd-Thomas, H., 2000. Score normalization for text-independent speaker verification systems. Digital Signal Processing 10 (1-3), 42–54.

Baker, B., Vogt, R., McLaren, M., Sridharan, S., 2009. Scatter difference NAP for SVM speaker recognition. In: Advances in Biometrics: Third International Conferences, ICB 2009, Alghero, Italy, June 2–5, 2009, Proceedings, 5558, p. 464.

Brummer, N., 2005. Focal: tools for fusion and calibration of automatic speaker detection systems. http://www.dsp.sun.ac.za/nbrummer/focal

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

Cheng, Z., Shen, B., Fan, X., Zhang, Y., 2008. Automatic coefficient selection in weighted maximum margin criterion. In: 19th International Conference on Pattern Recognition, 2008. ICPR 2008. IEEE, pp. 1–4.

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

Dehak, N., Dehak, R., Kenny, P., Brummer, N., Ouellet, P., Dumouchel, P., 2009. Support vector machines versus fast scoring in the low-dimensional total variability space for speaker verification. In: Proceedings of Interspeech, pp. 1559–1562.

Dehak, N., Kenny, P., Dehak, R., Dumouchel, P., Ouellet, P., 2010. Front-end factor analysis for speaker verification. IEEE Transactions on Audio, Speech, and Language Processing, PP (99), 1-1.

Garcia-Romero, D., Espy-Wilson, C., 2011. Analysis of i-vector length normalization in speaker recognition systems. In: International Conference on Speech Communication and Technology, pp. 249–252.

Hu, R., Jia, W., Huang, D., Lei, Y., 2010. Maximum margin criterion with tensor representation. Neurocomputing 73 (10), 1541–1549.

Kanagasundaram, A., Dean, D., Vogt, R., McLaren, M., Sridharan, S., Mason, M., 2012. Weighted LDA techniques for i-vector based speaker verification. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 4781–4784.

Kenny, P., 2005. Joint factor analysis of speaker and session variability: theory and algorithms. Tech. Rep., CRIM.

Kenny, P., 2010. Bayesian speaker verification with heavy tailed priors. In: Proceedings of Odyssey Speaker and Language Recognition Workshop, Brno, Czech Republic.

Kenny, P., Ouellet, P., Dehak, N., Gupta, V., Dumouchel, P., 2008. A study of inter-speaker variability in speaker verification. IEEE Transactions on Audio, Speech, and Language Processing 16 (5), 980–988.

Liang, Y., Li, C., Gong, W., Pan, Y., 2007. Uncorrelated linear discriminant analysis based on weighted pairwise fisher criterion. Pattern Recognition 40 (12), 3606–3615.

Loog, M., Duin, R., Haeb-Umbach, R., 2001. Multiclass linear dimension reduction by weighted pairwise Fisher criteria. IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (7), 762–766.

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

McLaren, M., van Leeuwen, D., 2011. Improved speaker recognition when using i-vectors from multiple speech sources. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5460–5463.

McLaren, M., van Leeuwen, D., 2011. Source-normalised-and-weighted LDA for robust speaker recognition using i-vectors. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 5456–5459.

McLaren, M., van Leeuwen, D., 2012. Source-normalized LDA for robust speaker recognition using i-vectors from multiple speech sources. IEEE Transactions on Audio, Speech, and Language Processing 20 (3), 755–766.

NIST, 2008. The NIST year 2008 speaker recognition evaluation plan. Tech. Rep., NIST. http://www.itl.nist.gov/iad/mig/tests/sre/2008/

NIST, 2010. The NIST year 2010 speaker recognition evaluation plan. Tech. Rep., NIST. http://www.itl.nist.gov/iad/mig/tests/sre/2010/

Pelecanos, J., Sridharan, S., 2001. Feature warping for robust speaker verification. In: 2001: A Speaker Odyssey – The Speaker Recognition Workshop, ISCA.

Price, J., Gee, T., 2005. Face recognition using direct, weighted linear discriminant analysis and modular subspaces. Pattern Recognition 38 (2), 209–219.

Prince, S.J., Elder, J.H., 2007. Probabilistic linear discriminant analysis for inferences about identity. In: IEEE 11th International Conference on Computer Vision, 2007. ICCV 2007. IEEE, pp. 1–8.

Senoussaoui, M., Kenny, P., Brummer, N., de Villiers, E., Dumouchel, P., 2011. Mixture of PLDA models in i-vector space for gender independent speaker recognition. In: Proceedings of INTERSPEECH, pp. 25–28.