WEIGHTED LDA TECHNIQUES FOR I-VECTOR BASED SPEAKER VERIFICATION

A. Kanagasundaram[1]*, D. Dean*[1]*, R. Vogt*[1]*, M. McLaren*[2]*, S. Sridharan*[1]*, M. Mason*[1]

[1] Speech Research Laboratory, Queensland University of Technology, Australia
[2] Centre for Language and Speech Technology, Radboud University Nijmegen, The Netherlands
[1]{a.kanagasundaram, r.vogt, d.dean, s.sridharan, m.mason}@qut.edu.au, [2]m.mclaren@let.ru.nl

ABSTRACT

This paper introduces the Weighted Linear Discriminant Analysis (WLDA) technique, based upon the weighted pairwise Fisher criterion, for the purposes of improving i-vector speaker verification in the presence of high inter-session variability. By taking advantage of the speaker discriminative information that is available in the distances between pairs of speakers clustered in the development i-vector space, the WLDA technique is shown to provide an improvement in speaker verification performance over traditional Linear Discriminant Analysis (LDA) approaches. A similar approach is also taken to extend the recently developed Source Normalised LDA (SNLDA) into Weighted SNLDA (WSNLDA) which, similarly, shows an improvement in speaker verification performance in both matched and mismatched enrolment/verification conditions. Based upon the results presented within this paper using the NIST 2008 Speaker Recognition Evaluation dataset, we believe that both WLDA and WSNLDA are viable as replacement techniques to improve the performance of LDA and SNLDA-based i-vector speaker verification.

Index Terms— speaker verification, i-vector, linear discriminant analysis

1. INTRODUCTION

Recent research in speaker verification has focused on the i-vector front-end factor analysis technique. This technique was first proposed by Dehak *et al.* [1] to provide an intermediate speaker representation between the high dimensional Gaussian Mixture Model (GMM) supervector and traditional low dimensional feature representations. The extraction of these intermediate-sized vectors, or i-vectors, were motivated by the existing supervector-based Joint Factor Analysis (JFA) approach, but rather than modelling the speaker and channel variability space separately, i-vectors are formed by modelling a single low-dimensional total-variability space that covers both speaker and channel variability [2]. Because channel variability is not explicitly removed in the i-vector extraction approach, channel compensation techniques must be implemented to limit the effects of channel variability in the i-vector speaker representations.

While the choice of channel compensation techniques is very much an active area of research, the use of Linear Discriminant Analysis (LDA) followed by Within Class Covariance Normalization (WCCN) used by Dehak *et al.* [2] has shown good performance. More recently, this approach was extended by McLaren and van Leeuwen [3] by proposing a new LDA-based approach, Source-Normalized LDA (SNLDA), which improve the i-vector-based speaker recognition in both mismatched conditions and conditions for which limited system development speech resources are available.

In this paper, we propose to investigate a new LDA technique, based upon the weighted pair-wise Fisher criteria [4], that has recently shown promise in the field of template-based face recognition [5]. This technique, known as Weighted LDA (WLDA), takes advantage of the discriminatory information between pairs of classes, or speakers for our application, in the between-class scatter that has not yet been investigated for i-vector-based speaker verification. By applying a weighted parameter to class pairs that weights closer pairs higher, WLDA should provide an improvement in discriminative ability between classes that would otherwise be difficult to distinguish in the LDA- or SNLDA-transformed i-vector space. Motivated by the improvements obtained for WLDA over traditional LDA for face recognition [5], our aim in this paper is to investigate if a similar approach can be taken with WLDA and Weighted SNLDA (WSNLDA) to provide improvements for i-vector-based speaker verification.

This paper is structured as follows. Section 2 gives a brief introduction to process of i-vector based speaker verification system. Section 3 details the proposed W-LDA and W-SNLDA techniques. The experimental protocol and corresponding results are given in Section 4 and Section 5.

2. SPEAKER VERIFICATION USING I-VECTORS

In contrast to the separate speaker and channel dependent subspaces of JFA, i-vectors represent the GMM supervector using a single total-variability subspace. 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 [6]. An i-vector speaker and channel dependent GMM supervector can be represented by

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

where $\mathbf{m}$ is the speaker and channel independent background UBM supervector, $\mathbf{T}$ is a low rank matrix representing the primary directions variation across a large collection of development data. Finally, $\mathbf{w}$ is normally distributed with parameters $N(0,1)$, and is the *i-vector* representation used for speaker verification.

For details of the total variability subspace training and subsequent i-vector extraction, the reader is encouraged to investigate the techniques covered by Dehak *et al.* [2].

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