Available online at www.sciencedirect.com
ScienceDirect

Computer Speech and Language 00 (2013) 1–23

~~Computer~~
Speech
and
~~Language~~
www.elsevier.com/locate/procedia

I-vector based Speaker Recognition Using Advanced Channel Compensation Techniques

Ahilan Kanagasundaram⋆, David Dean⋆, Sridha Sridharan⋆, Mitchell McLaren†, Robbie Vogt⋆

⋆ *Speech and Audio Research Lab, SAIVT, Queensland University of Technology, Australia.*
† *Centre for Language and Speech Technology, Radboud University Nijmegen, The Netherlands.*
a.kanagasundaram@qut.edu.au, d.dean@qut.edu.au, s.sridharan@qut.edu.au,
m.mclaren@let.ru.nl, r.vogt@qut.edu.au

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.* [8] to provide an intermediate speaker representation between the high-dimensional Gaussian mixture model (GMM) super-vector and traditional low-dimensional mel-frequency cepstral coefficients (MFCC) 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 [12, 14]. 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 [6]. This approach was motivated by Dehak *et al.* finding that i-vectors don't lose any speaker discriminant information, unlike the JFA approach, where some speaker discriminant information is lost in the channel space [6]. As the channel variability is included within the