Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 00 (2014) 1–18

Speech communication

www.elsevier.com/locate/procedia

Improving Short Utterance I-vector Speaker Verification using Utterance Variance Modelling and Compensation Techniques

A.Kanagasundaram⋆, D.Dean⋆, S.Sridharan⋆, J.Gonzalez-Dominguez†, J.Gonzalez-Rodriguez†, D.Ramos†

⋆ *Speech Research Lab, SAIVT, Queensland University of Technology, Australia.*
† *ATVS Biometric Recognition Group, Universidad Autonoma de Madrid, Spain.*

a.kanagasundaram@qut.edu.au, d.dean@qut.edu.au, s.sridharan@qut.edu.au, javier.gonzalez@uam.es, joaquin.gonzalez@uam.es, daniel.ramos@uam.es

Abstract

This paper proposes techniques to improve the performance of i-vector based speaker verification system when only short utterances are available. Short-length utterance i-vectors vary with speaker, session variations, and the phonetic content of the utterance. Well established methods such as linear discriminant analysis (LDA), source-normalized LDA (SN-LDA) and within-class covariance normalisation (WCCN) exist for compensating the session variation but we have identified the variability introduced by phonetic content due to utterance variation as an additional source of degradation when short-duration utterances are used. To compensate for utterance variations in short i-vector based speaker verification systems using cosine similarity scoring (CSS), we have introduced a short utterance variance normalization (SUVN) technique and a short utterance variance (SUV) modelling approach at the i-vector feature level. A combination of SUVN with LDA and SN-LDA is proposed to compensate the session and utterance variations and is shown to provide improvement in performance over the traditional approach of using LDA and/or SN-LDA followed by WCCN. An alternative approach is also introduced using the probabilistic linear discriminant analysis (PLDA) approach to directly model the SUV. The combination of SUVN, LDA and SN-LDA followed by SUV PLDA modelling provides an improvement over the baseline PLDA approach. We also show that for this combination of techniques, the utterance variation information needs to be artificially added to full-length i-vectors for PLDA modelling.

Keywords: Speaker verification, I-vector, PLDA, SN-LDA, SUVN, SUV

1. Introduction

Many remarkable advances in dealing with mismatch between enrolment and verification for speaker verification have been accomplished during the last few years, which have led to highly reliable performance when reasonable amounts of speech are available. Techniques based on factor analysis, such as joint factor analysis (JFA) [9, 23], i-vectors [3] and probabilistic linear discriminant analysis (PLDA) [10], have demonstrated outstanding behaviour in challenging evaluation scenarios, such as the speaker recognition evaluation series developed by the National Institute of Standards and Technology (NIST) [18, 19]. Unfortunately, the performance of many of these approaches degrades rapidly as the available amount of enrolment and/or verification speech decreases [21, 7, 8], limiting the utility of speaker verification in real world applications, such as access control or forensics.

The total-variability, or i-vector, approach has risen to prominence as the de-facto standard in recent state-of-the-art speaker verification systems, due to its intrinsic capability to map an utterance to a single