# IMPROVING OUT-DOMAIN PLDA SPEAKER VERIFICATION USING UNSUPERVISED INTER-DATASET VARIABILITY COMPENSATION APPROACH

*Ahilan Kanagasundaram, David Dean and Sridha Sridharan*

Speech Research Laboratory, Queensland University of Technology, Australia

{a.kanagasundaram, d.dean, s.sridharan}@qut.edu.au

**ABSTRACT**

Experimental studies have found that when the state-of-the-art probabilistic linear discriminant analysis (PLDA) speaker verification systems are trained using out-domain data, it significantly affects speaker verification performance due to the mismatch between development data and evaluation data. To overcome this problem we propose a novel unsupervised inter dataset variability (IDV) compensation approach to compensate the dataset mismatch. IDV-compensated PLDA system achieves over 10% relative improvement in EER values over out-domain PLDA system by effectively compensating the mismatch between in-domain and out-domain data.

***Index Terms*—** speaker verification, PLDA, domain adaptation, inter-dataset variability

## 1. INTRODUCTION

A significant amount of development data, especially in the presence of large intersession variability, is required to develop a speaker verification system. Recent studies have found that when speaker verification is developed in one domain data and evaluated in another domain data, the dataset mismatch significantly affects the speaker verification performance [1, 2, 3]. Therefore significant amount of target domain data is required to develop speaker verification system in order to achieve state-of-the-art performance. However, it is hard to collect adequate amount of target domain data, specially speaker labelled data in real world environments. In recent times, researchers have been proposing several approaches to achieve state-of-the-art speaker verification performance if significant amount of out-domain data and limited in-domain unlabelled is available. This problem is defined as domain adaptation.

Recently, Garcia-Romero *et at.* [1] have found that the adaptation of the PLDA parameters produces the largest gains, and universal background model (UBM) and total-variability matrix would not be required to estimate on in-domain data. They have studied several supervised approaches, including fully Bayesian adaptation, approximate *maximum a posteriori* (MAP) adaptation, weighted likelihood [1]. Aronowitz [2] introduced inter dataset variability compensation (IDVC) to explicitly compensate for dataset shift in the i-vector space, which is based on nuisance attribute projection (NAP) method. For IDVC estimation, out-domain Switchboard dataset is partitioned into several sub datasets. Recently, Garcia-Romero *et at.* [4] have also introduced agglomerative hierarchical clustering (AHC) based unsupervised approach for domain adaptation.

In this paper, a novel unsupervised inter-dataset variability (IDV) is introduced in order to compensate the mismatch between out-domain data and in-domain data. Our approach is similar to the IDVC approach proposed by Aronowitz in [2] but in contrast, out-domain Switchboard dataset is not required to be partition into several subsets to estimate the inter dataset variability compensation. Recently, we have proposed short utterance variance (SUV) approach to capture the utterance variation for short utterance PLDA speaker verification system [5, 6]. In this paper, similar idea is taken to capture the variation between in-domain and out-domain data. The variation between in-domain and out-domain data is defined as the outer product of difference between out-domain i-vectors and average of in-domain i-vectors. The IDV compensation matrix is estimated using Cholesky decomposition of inverse of variation matrix. We analyse how the limited in-domain unlabelled data that is available for IDV compensation estimation, affects the speaker verification performance.

This paper is structured as follows: Section 2 details the i-vector feature extraction techniques. Section 3 details the inter dataset variability compensation approach. Section 4 explains the Gaussian PLDA (GPLDA) based speaker verification system. The experimental protocol and corresponding results are given in Section 5 and Section 6. Section 7 concludes the paper.

## 2. I-VECTOR FEATURE EXTRACTION

I-vectors represent the Gaussian mixture model (GMM) super-vector by 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 [7]. An i-vector speaker and channel dependent GMM super-vector can be