Int J Speech Technol (2017) 20:247–259
DOI 10.1007/s10772-017-9402-8

A study on the effects of using short utterance length development data in the design of GPLDA speaker verification systems

Ahilan Kanagasundaram[1,2] · David Dean[2] · Sridha Sridharan[2] · Houman Ghaemmaghami[2] · Clinton Fookes[2]

Received: 18 October 2016 / Accepted: 25 January 2017 / Published online: 16 February 2017

Abstract This paper studies the performance degradation of Gaussian probabilistic linear discriminant analysis (GPLDA) speaker verification system, when only short-utterance data is used for speaker verification system development. Subsequently, a number of techniques, including utterance partitioning and source-normalised weighted linear discriminant analysis (SN-WLDA) projections are introduced to improve the speaker verification performance in such conditions. Experimental studies have found that when short utterance data is available for speaker verification development, GPLDA system overall achieves best performance with a lower number of universal background model (UBM) components. As a lower number of UBM components significantly reduces the computational complexity of speaker verification system, that is a useful observation. In limited session data conditions, we propose a simple utterance-partitioning technique, which when applied to the LDA-projected GPLDA system shows over 8% relative improvement on EER values over baseline system on NIST 2008 truncated 10–10 s conditions. We conjecture that this improvement arises from the apparent increase in the number of sessions arising from our partitioning technique and this helps to better model the GPLDA parameters. Further, partitioning SN-WLDA-projected GPLDA shows over 16% and 6% relative improvement on EER values over LDA-projected GPLDA systems respectively on NIST 2008 truncated 10–10 s *interview-interview*, and NIST 2010 truncated 10–10 s *interview-interview* and *telephone-telephone* conditions.

✉ Ahilan Kanagasundaram
ahilan@eng.jfn.ac.lk

David Dean
d.dean@qut.edu.au

Sridha Sridharan
s.sridharan@qut.edu.au

Houman Ghaemmaghami
houman.ghaemmaghami@qut.edu.au

Clinton Fookes
c.fookes@qut.edu.au

[1] Department of Electrical & Electronic Engineering, Faculty of Engineering, University of Jaffna, Kilinochchi, Sri Lanka

[2] Speech Research Lab, SAIVT, Queensland University of Technology, Brisbane, QLD, Australia

1 Introduction

Speaker verification has traditionally required a large volume of speech from a large number of different speakers for reliable development and evaluation, particularly in the presence of high intersession variability. However, it can be hard to acquire a sufficient duration of speech data and a larger number of speaker session data in many real-world environments, limiting the suitability of speaker verification for many everyday applications. Recently, a number of interesting techniques have been focusing on reducing the volume of speech required to develop new models for deployment into previously unseen environments (Kanagasundaram et al. 2012b, 2013a, 2014a; Vogt et al. 2008a; Kenny et al. 2013; McLaren et al. 2010). Reducing the amount of speech required during enrolment and verification whilst maintaining satisfactory performance has been the focus in a number of recent studies across a range of speaker verification techniques: joint factor analysis (JFA) (Vogt et al. 2008), support vector machines (SVM) (McLaren et al. 2010), i-vectors (Kanagasundaram et al. 2011) and probabilistic linear discriminant analysis (PLDA) (Kanagasundaram et al. 2012b, 2014b). These