Modelling and Compensation for Language Mismatch in Speaker Verification

Abhinav Misra, John H.L. Hansen

PII: S0167-6393(17)30044-4
DOI: 10.1016/j.specom.2017.09.004
Reference: SPECOM 2492

To appear in: *Speech Communication*

Received date: 22 January 2017
Revised date: 26 September 2017
Accepted date: 28 September 2017

Please cite this article as: Abhinav Misra, John H.L. Hansen, Modelling and Compensation for Language Mismatch in Speaker Verification, *Speech Communication* (2017), doi: 10.1016/j.specom.2017.09.004

Modelling and Compensation for Language Mismatch in Speaker Verification

Abhinav Misra, John H. L. Hansen[1]

Center for Robust Speech Systems (CRSS)
Erik Jonsson School of Engineering & Computer Science
The University of Texas at Dallas (UTD), Richardson, Texas, USA
{abhinav.misra, john.hansen}@utdallas.edu

Abstract

Language mismatch represents one of the more difficult challenges in achieving effective speaker verification in naturalistic audio streams. The portion of bi-lingual speakers worldwide continues to grow making speaker verification for speech technology more difficult. In this study, three specific methods are proposed to address this issue. Experiments are conducted on the PRISM (Promoting Robustness in Speaker Modeling) evaluation-set. We first show that adding small amounts of multi-lingual seed data to the Probabilistic Linear Discriminant Analysis (PLDA) development set, leads to a significant relative improvement of +17.96% in system Equal Error Rate (EER). Second, we compute the eigendirections that represent the distribution of multi-lingual data added to PLDA. We show that by adding these new eigendirections as part of the Linear Discriminant Analysis (LDA), and then minimizing them to directly compensate for language mismatch, further performance gains for speaker verification are achieved. By combining both multi-lingual PLDA and this minimization step with the new set of eigendirections, we obtain a +26.03% relative improvement in EER. In practical scenarios, it is highly unlikely that multi-lingual seed data representing the languages present in the test-set would be

[1]This project was funded by AFRL under contract FA8 750-12-1-0188 and partially by the University of Texas at Dallas from the Distinguished University Chair in Telecommunications Engineering held by J. H. L. Hansen.