Odyssey 2012
The Speaker and Language Recognition Workshop
25-28 June 2012, Singapore

# PLDA based Speaker Recognition on Short Utterances

Ahilan Kanagasundaram, Robbie Vogt, David Dean, Sridha Sridharan

Speech and Audio Research Laboratory
Queensland University of Technology, Brisbane, Australia
{a.kanagasundaram, r.vogt, d.dean, s.sridharan}@qut.edu.au

## Abstract

This paper investigates the effects of limited speech data in the context of speaker verification using a probabilistic linear discriminant analysis (PLDA) approach. Being able to reduce the length of required speech data is important to the development of automatic speaker verification system in real world applications. When sufficient speech is available, previous research has shown that heavy-tailed PLDA (HTPLDA) modeling of speakers in the i-vector space provides state-of-the-art performance, however, the robustness of HTPLDA to the limited speech resources in development, enrolment and verification is an important issue that has not yet been investigated. In this paper, we analyze the speaker verification performance with regards to the duration of utterances used for both speaker evaluation (enrolment and verification) and score normalization and PLDA modeling during development. Two different approaches to total-variability representation are analyzed within the PLDA approach to show improved performance in short-utterance mismatched evaluation conditions and conditions for which insufficient speech resources are available for adequate system development.

The results presented within this paper using the NIST 2008 Speaker Recognition Evaluation dataset suggest that the HTPLDA system can continue to achieve better performance than Gaussian PLDA (GPLDA) as evaluation utterance lengths are decreased. We also highlight the importance of matching durations for score normalization and PLDA modeling to the expected evaluation conditions. Finally, we found that a pooled total-variability approach to PLDA modeling can achieve better performance than the traditional concatenated total-variability approach for short utterances in mismatched evaluation conditions and conditions for which insufficient speech resources are available for adequate system development.

## 1. Introduction

In a typical speaker verification system, the significant amount of speech required for reliable speaker evaluations (enrolment and verification) in the presence of large inter-session variability has limited the widespread use of speaker verification technology in everyday applications. Reducing the amount of speech required while obtaining satisfactory performance has been the focus of a number of recent studies in state-of-the-art speaker verification design, including joint factor analysis (JFA), i-vectors and support vector machines (SVM). These studies have shown that performance degrades considerably in very short utterances ($< 10s$) for all common approaches [1, 2, 3, 4]. This paper will focus on whether a recently proposed probabilistic linear discriminant analysis (PLDA) approach to speaker verification could form a suitable foundation for continuing research into short utterance speaker verification.

The PLDA technique was originally proposed by Price *et al.* [5] for face recognition, and later adapted for modeling i-vector distributions for speaker verification by Kenny *et al.* [6, 7, 8]. In his initial work, Kenny investigated two PLDA approaches, Gaussian PLDA (GPLDA) and heavy-tailed PLDA (HTPLDA) [6]. For GPLDA, the speaker and channel subspaces are modeled with Gaussian distributions, but a major limitation of this approach is the lack of robustness to outliers in the speaker and channel subspaces [6]. In order to better cope with these outliers, Kenny proposed that Student's t-distribution can be used as an alternative for model the subspaces as an alternative to the Gaussian. As Student's t-distribution has heavier tails compared to the exponentially-decaying tails of a Gaussian, this approach provides a better representation of the full subspace, including the outliers [9]. Kenny has found that both PLDA approaches, and in particular, HTPLDA achieved significant improvement over JFA on the standard NIST SRE conditions [6], but these approaches have not yet been investigated under short utterance evaluation and development data conditions.

The main aim of this paper is to investigate the effect of only having short utterances available for evaluation and development for the PLDA speaker verification. The i-vector subspace has been shown to provide a more speaker discriminative representation than JFA factor analysis, and we believe that the PLDA modeling approach will continue to work well, in comparison to other approaches, as the length of evaluation utterances are decreased. In this paper, we will closely investigate the performance of Guassian and heavy-tailed PLDA in short utterance evaluation and development data conditions in order to investigate the best approach available for limited data speaker verification. As well as the matched *telephone-telephone* enrolment-verification conditions, we will also investigate the impact of mismatched and matched *interview-interview* conditions, using two approaches to combining the mismatched conditions in the i-vector total variability representation, in short utterance PLDA speaker verification.

This paper is structured as follows. Section 2 gives a brief introduction to PLDA based speaker verification system. The experimental protocol and corresponding results are given in Section 3 and Section 4. The paper is concluded in Section 5.

## 2. Speaker verification using PLDA techniques

PLDA is a generative model which was adapted from face recognition to model i-vector distributions for speaker verification by Kenny [6]. This approach can be seen to be very similar to the JFA approach, but using i-vectors rather than Gaussian mixture model (GMM) super-vectors as the basis for factor modeling.

### 2.1. I-vector feature extraction

I-vectors represent the 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 [10]. An i-vector speaker and channel dependent GMM super-vector can be represented by,

$$\boldsymbol{\mu} \quad = \quad \mathbf{m} + \mathbf{Tw}, \tag{1}$$

where **m** is the same universal background model (UBM) super-vector used in the JFA approach and **T** is a low rank total-variability matrix. It is assumed that i-vectors (**w**) are normally distributed with parameters *N(0,1)*. Extracting an i-vector from the total-variability subspace is essentially a *maximum a-posteriori adaptation* (MAP) of **w** in the subspace defined by **T**. An efficient procedure for the optimization of the total-variability subspace **T** and subsequent extraction of i-vectors is described in [11] and [12].

The total-variability subspace is responsible for defining a suitable subspace from which i-vectors are extracted. Telephone-speech speaker verification is investigated in Section 4.1, where the total-variability subspace (${R_w}^{tel} = 500$) is trained from telephone speech development data. Combined telephone and microphone speaker verification is investigated in Section 4.2, and for this approach the total-variability subspace should be trained in a manner that best exploits the useful speaker variability contained in speech acquired from both telephone and microphone sources. McLaren and van Leeuwen have investigated different types of total-variability representations with i-vector systems [13], and they found that a pooled total-variability approach provides a better representation for i-vector speaker verification when compared to the traditional concatenated approach. In this paper, both the pooled and concatenated total-variability approach will be investigated for PLDA speaker verification. For the pooled total-variability approach, the total-variability subspace (${R_w}^{telmic} = 500$) is trained on telephone and microphone speech utterances together. For the concatenated total-variability approach, the separate total-variability telephone-only subspace (${R_w}^{tel} = 400$) and microphone-only subspace (${R_w}^{mic} = 100$) are trained separately using telephone and microphone speech, then both subspace transformations are concatenated to create a single total-variability space.

### 2.2. PLDA modeling

Rather than attempting to compensate for intersession variability in the i-vector space, a more sophisticated attempt to directly model session and speaker variability within the i-vector space was recently proposed by Kenny [6] as PLDA. A speaker and channel dependent i-vector, **w** can be defined as

$$\mathbf{w}_r \quad = \quad \bar{\mathbf{w}} + \mathbf{U}_1\mathbf{x}_1 + \mathbf{U}_2\mathbf{x}_{2r} + \boldsymbol{\varepsilon}_r \tag{2}$$

where for given speaker recordings $r = 1, .....R$; $\mathbf{U}_1$ is the eigenvoice matrix and $\mathbf{U}_2$ is the eigenchannel matrix, $\mathbf{x}_1$ and $\mathbf{x}_{2r}$ are the speaker and channel factors respectively and $\boldsymbol{\varepsilon}_r$ is the residuals. In the PLDA modeling, the speaker specific part can be represented as $\bar{\mathbf{w}} + \mathbf{U}_1\mathbf{x}_1$, which represents the between speaker variability. The covariance matrix of the speaker part is $\mathbf{U}_1{\mathbf{U}_1}^T$. The channel specific part is represented as $\mathbf{U}_2\mathbf{x}_{2r} + \boldsymbol{\varepsilon}_r$, which describes the within speaker variability. The covariance matrix of channel part is $\boldsymbol{\Lambda}^{-1} + \mathbf{U}_2{\mathbf{U}_2}^T$. We assume that precision matrix ($\boldsymbol{\Lambda}$) is full rank and remove the eigenchannel ($\mathbf{U}_2$) from equation (2), as we found that PLDA speaker verification didn't show major improvement with eigenchannels, and removing them provided a useful decrease in computational complexity.

#### *2.2.1. GPLDA*

In GPLDA, we assume that speaker factors ($\mathbf{x}_1$) have a standard normal distribution of dimension $N_1$, and the residuals ($\boldsymbol{\varepsilon}_r$) also have a standard normal distribution with mean 0 and a covariance matrix ($\boldsymbol{\Lambda}^{-1}$). In GPLDA, the model parameters, **m**, $\mathbf{U}_1$, and $\boldsymbol{\Lambda}$ are estimated from development i-vectors. Because of outliers in the i-vectors space, the choice of Gaussian for modeling in GPLDA is not optimal and this led to the development of the HTPLDA approach.

#### *2.2.2. HTPLDA*

For the HTPLDA approach, Kenny proposed using Student's t-distribution for modeling the speaker and channel subspaces as an alternative to the Gaussian distribution of GPLDA [6]. In this approach, we assume that speaker factors and residual factors can be modeled by a heavy-tailed distribution to provide better representation of the outliers in the i-vector space. These speaker and residual factors are scaled by gamma distribution scalars, which can be represented as follows,

$$\mathbf{x}_1 \sim N(0, {u_1}^{-1}I) \text{ where } u_1 \sim G(n_1/2, n_1/2)$$
$$\boldsymbol{\varepsilon}_r \sim N(0, {\upsilon_r}^{-1}\boldsymbol{\Lambda}^{-1}) \text{ where } \upsilon_1 \sim G(\nu/2, \nu/2)$$

where $n_1$ and $\nu$ are degrees of freedom, $u_1$, $\upsilon_r$ are gamma distribution scalers, $N(\mu, \Sigma)$ is a Gaussian distribution with mean $\mu$ and covariance $\Sigma$, and $G$(a,b) is a gamma distribution with shape parameter $a$ and scale parameter $b$. In HTPLDA, the model parameters, **m**, $\mathbf{U}_1$, $\boldsymbol{\Lambda}$, $n_1$, and $\nu$ are estimated from the development i-vectors.

#### *2.2.3. PLDA scoring*

For PLDA, scoring is conducted using the batch-likelihood ratio between a target and test i-vector [6]. Given two i-vectors, $w_{target}$ and $w_{test}$, the batch likelihood ratio can be calculated as follows,

$$\ln \frac{P(w_{target}, w_{test} \mid H_1)}{P(w_{target} \mid H_0)P(w_{test} \mid H_0)} \tag{3}$$

where $H_1$ denotes the hypothesis that the i-vectors represent the same speaker and $H_0$ denotes that they do not.

## 3. Experimental configuration

The PLDA experiments were evaluated using the NIST 2008 Speaker Recognition Evaluation (SRE) utterances from the *short2-short3* and *10sec-10sec* evaluation conditions. The shortened utterances were obtained by truncating the NIST2008 *short2-short3* conditions to the specified length of active speech for both enrolment and verification. Prior to the truncation, the first 20 seconds of active speech were removed from all utterances to avoid capturing similar introductory statements across multiple utterances. The performance was evaluated using the equal error rate (EER) and the minimum decision cost function (DCF), calculated using $C_{miss} = 10$, $C_{FA} = 1$, and $P_{target} = 0.01$. In order to evaluate the PLDA approaches in both matched and mismatched conditions, evaluation was performed using the NIST 2008 *telephone-telephone*, *interview-interview*, *telephone-interview* and

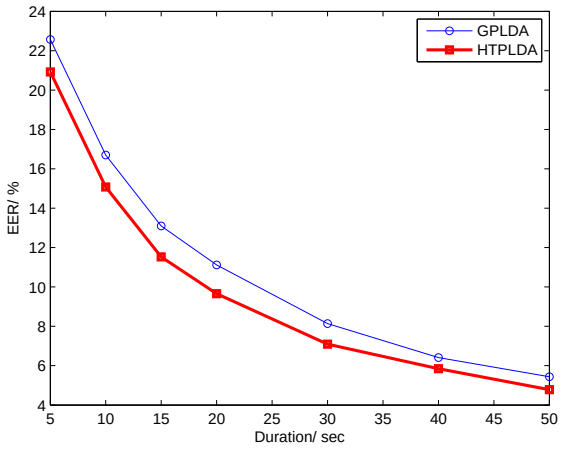
Table 1: *Comparison of GPLDA and HTPLDA systems with and without S-Norm on the common set of the 2008 NIST SRE standard conditions. (a) GPLDA (b) HTPLDA. The best performing systems by both EER and DCF are highlighted across each row.*

(a) *GPLDA*

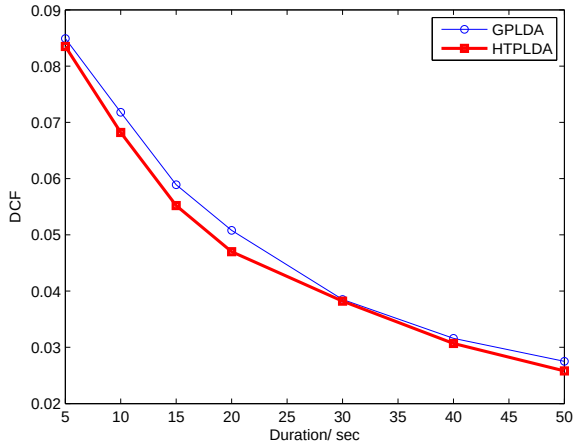
| Evaluation | Without Snorm | | With Snorm | |
|---|---|---|---|---|
| utterance lengths | EER | DCF | EER | DCF |
| short2-short2 | 4.20% | 0.0204 | **3.13%** | **0.0163** |
| 10sec-10sec | 19.94% | 0.0837 | **15.23%** | **0.0690** |

(b) *HTPLDA*

| Evaluation | Without Snorm | | With Snorm | |
|---|---|---|---|---|
| utterance lengths | EER | DCF | EER | DCF |
| short2-short3 | **2.39%** | **0.0128** | 2.47% | 0.0151 |
| 10sec-10sec | 16.14% | 0.0741 | **13.89%** | **0.0649** |




(a) EER




(b) DCF

Figure 1: *Comparison of GPLDA and HTPLDA systems at different lengths of active speech for each enrolment and verification condition*, (a) EER, *and* (b) DCF

*interview-telephone* enrolment-verification conditions. Evaluation was performed using only the English evaluation conditions.

We used 13 dimensional feature-warped mel frequency cepstral coefficients (MFCC) with appended delta coefficients. Two gender-dependent UBMs containing 512 Gaussians trained on NIST 2004 SRE corpus are used throughout our experiments. These gender-dependent UBMs were used to calculate the Baum-Welch statistics for calculation of the total-variability subspace of dimension $R_w = 500$, which is then used to calculate the i-vector speaker representations.

For the initial telephone speech speaker verification experiments, the development data for the total-variability subspace and the PLDA modeling were obtained from the telephone-only utterances available in NIST 2004, 2005 and 2006 SRE corpora as well as Switchboard II. We empirically selected 90 eigenvoices ($N_1$) based upon speaker verification performance, and the precision matrix ($\mathbf{\Lambda}$) was defined as full rather than diagonal. For Snorm, the statistics were calculated using telephone-only utterances from the NIST04, 05 corpora [14].

When we introduce the mismatched and *interview-interview* evaluation environments, both pooled and concatenated total-variability spaces are calculated across both the telephone and microphone data available in the same development datasets outlined above. For Snorm, the statistics were calculated using both telephone and microphone speech NIST04, 05 and 06 corpora. For the mismatched and matched *interview-interview* evaluation experiments, we empirally set the number of eigenvoices ($N_1$) to 100 based upon speaker verification performance, and kept the precision matrix full, as in the telephone experiments.

## 4. Results

Following is an experimental study regarding the impact of limited speech on PLDA speaker verification. Experiments studies are divided into two sections. Telephone speech based PLDA system is investigated with limited data conditions in the first section. Initially experiments look at NIST standard conditions before progressively investigating on short utterance evaluation and development data conditions. In the second section, telephone and microphone speech based PLDA system is investigated with NIST standard and truncated conditions.

### 4.1. Analysis of short utterance performance on telephone speech based PLDA system

Initially the GPLDA and HTPLDA systems were investigated with NIST standard evaluation conditions using only telephone utterances. Table 1 presents results comparing the performance of the GPLDA and HTPLDA systems with and without S-Norm on the standard NIST SRE 08 evaluation conditions. As had been previously shown by Kenny [6], we have confirmed that the HTPLDA system provides an improvement over GPLDA. Similarly to Kenny, we have found that S-Norm improves the performance of the GPLDA system in both the *short2-short3* and the *10sec-10sec* enrolment-verification conditions. These results also indicate that, while there appears to be limited disadvantage to score normalization in longer utterances, HTPLDA is improved by score normalisation for short utterances.

In order to more closely examine the behavior of PLDA speaker verification for short utterances, we evaluated both the GPLDA and HTPLDA systems for truncated evaluation data, as shown in the Figure 1. These results show that the HTPLDA

Table 2: *Performance comparison of GPLDA and HTPLDA systems with full and matched length score normalization data (a) GPLDA (b) HTPLDA. The best performing systems by both EER and DCF are highlighted across each row.*

(a) *GPLDA*

| Evaluation utterance lengths | S-Norm development data: Full length EER | Full length DCF | Matched length EER | Matched length DCF |
|---|---|---|---|---|
| 5 sec - 5 sec | 22.57% | **0.0849** | **22.32%** | 0.0855 |
| 10 sec - 10 sec | 16.70% | 0.0718 | **16.65%** | **0.0716** |
| 15 sec - 15 sec | 13.10% | 0.0589 | **12.52%** | **0.0587** |
| 20 sec - 20 sec | 11.12% | **0.0508** | **11.04%** | 0.0513 |

(b) *HTPLDA*

| Evaluation utterance lengths | S-Norm development data: Full length EER | Full length DCF | Matched length EER | Matched length DCF |
|---|---|---|---|---|
| 5 sec - 5 sec | 20.92% | 0.0835 | **20.76%** | **0.0828** |
| 10 sec - 10 sec | **15.08%** | **0.0682** | **15.08%** | 0.0692 |
| 15 sec - 15 sec | 11.53% | **0.0552** | **11.37%** | 0.0563 |
| 20 sec - 20 sec | 9.66% | **0.0470** | **9.55%** | 0.0480 |

Table 3: *Performance comparison of GPLDA systems with full and matched length PLDA modeling data, HTPLDA systems with full and mixed length PLDA modeling data (a) GPLDA (b) HTPLDA. The best performing systems by both EER and DCF are highlighted across each row.*

(a) *GPLDA*

| Evaluation utterance lengths | GPLDA development data: Full length EER | Full length DCF | Matched length EER | Matched length DCF |
|---|---|---|---|---|
| 10sec - 10sec | 16.70% | 0.0718 | **16.04%** | **0.0679** |
| 20sec - 20sec | 11.12% | 0.0508 | **10.63%** | **0.0490** |

(b) *HTPLDA*

| Evaluation utterance lengths | HTPLDA development data: Full length EER | Full length DCF | Mixed length EER | Mixed length DCF |
|---|---|---|---|---|
| 10sec - 10sec | 15.08% | 0.0682 | **13.67%** | **0.0639** |
| 20sec - 20sec | 9.66% | 0.0470 | **9.07%** | **0.0461** |

system continues to achieve better performance than GPLDA for all the truncated conditions, although the difference is not as dramatic for DCF as for EER. Overall, the results show that as the utterance length decreases, performance degrades at an increasing rate, rather than in proportion with the reduced length. From these results, we believe that HTPLDA provides a good choice for speaker verification in very short evaluation conditions.

Finally, the GPLDA and HTPLDA systems were analyzed with short utterance development data for both normalisation and PLDA modeling. Table 2 presents the results comparing the performance of the GPLDA and HTPLDA systems with full-length score normalization and matched-length score normalization (score normalization data truncated to same length as evaluation data). We found that matched-length score normalization improves the EER performance of both PLDA systems across all truncated conditions, but doesn't show consistent improvement of DCF. This shows, that rather than being a hindrance to normalisation performance, short utterance development data (if matched in length), can improve normalisation for speaker verification.

Secondly the GPLDA and HTPLDA systems were investigated with short utterance PLDA modeling development data. Table 3(a) presents the results of the GPLDA speaker verification system trained during development on full-length utterances and utterances with lengths matched to the evaluation conditions. These results suggest that when the GPLDA system is modeled with matched-length utterances, improvement can be achieved over modeling based upon full-length utterances.

When attempting to model the matched short utterances with HTPLDA, we found that we could not fit the i-vectors with a heavy-tailed distribution. Because of this difficulty and the improvement in GPLDA modeling with matched utterances, we believe that this is an indication that short utterances in the i-vector space have less outliers than full-length utterances, and therefore are better modeled with Gaussians.

In order to still be able to take advantage of matching the development data with evaluation, we attempted to model the short-utterance HTPLDA system by including both matched and full-length utterances in the development data. This approach is shown as the 'Mixed' column in Table 3(b). We can see that the mixed-length HTPLDA modeling provided improved speaker verification performance over the full-utterances modeling. We believe that while matching the i-vector lengths does not appear to be feasible in HTPLDA modeling, the mixed-length modeling approach provides a closer match between development and evaluation, providing for an improvement in speaker verification performance in short utterance evaluation conditions.

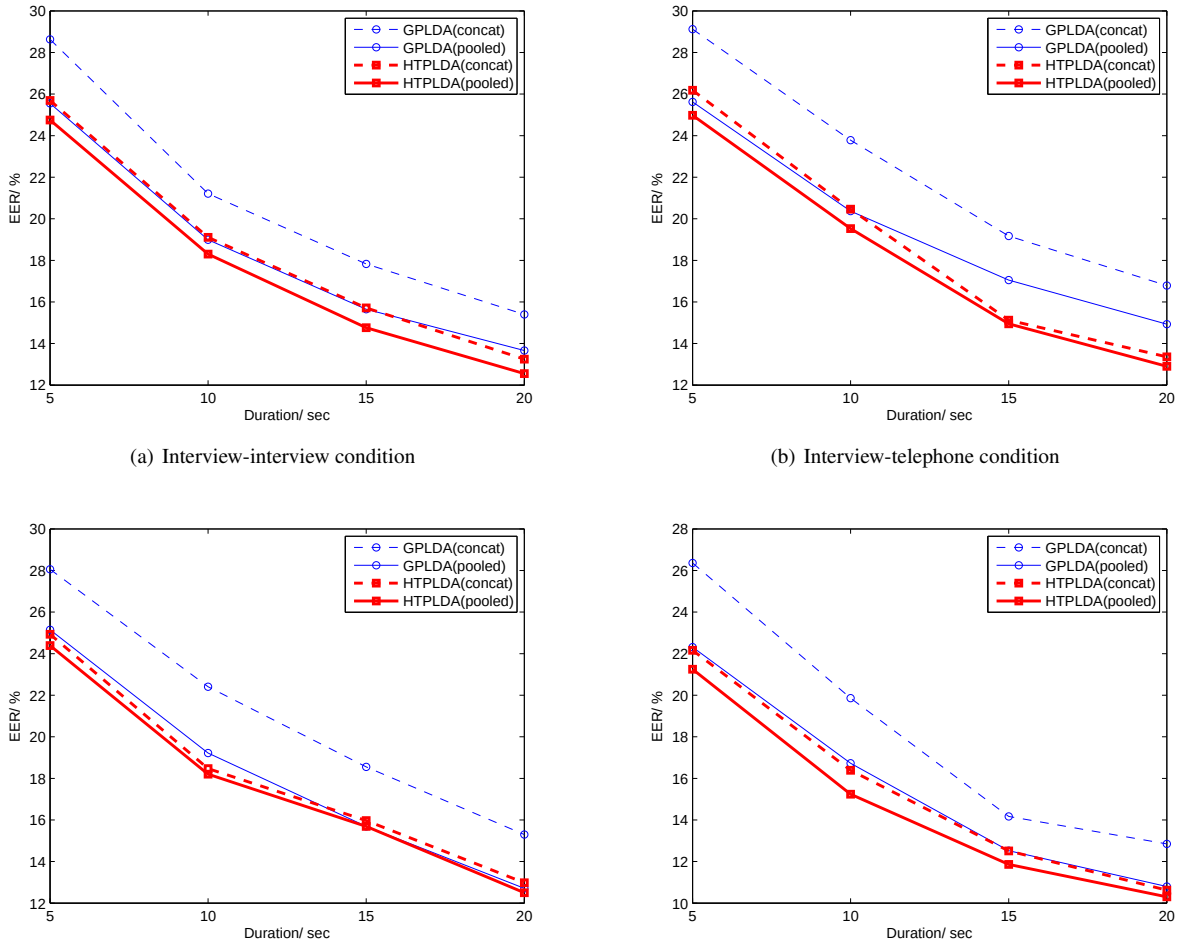
### 4.2. Analysis of short utterance performance on telephone and microphone speech based PLDA system

In the previous section, the PLDA approaches were investigate with solely telephone speech during both development and evaluation. In this section, we will expand our evaluation of PLDA approaches to mismatched and limited-data channel conditions with two different total-variability modeling approaches.

The EER performance of pooled and concatenated total-variability modeling for GPLDA and HTPLDA systems in short utterance evaluation data is shown in Figure 2. All results are presented with S-Norm applied. From the figure, it can be seen that the pooled total-variability approach provided improved performance for both the GPLDA and HTPLDA speaker verification systems across all lengths and channel conditions. These results also suggest that when the utterance length is reduced, the pooled total-variability approach improves the performance at increasing rate. It has also been found that the pooled total-variability approach achieves considerable improvement on *telephone-telephone* and *interview-interview* matched conditions across all truncated evaluation data for the HTPLDA system.

## 5. Conclusions

The challenges of providing robust speaker verification for applications with access to only short speech utterances remains a key hurdle to the broad adoption of speaker verification systems. This paper presented a study on the effects of limited speech data on PLDA based speaker verification.




(a) Interview-interview condition

(b) Interview-telephone condition

(c) Telephone-interview condition

(d) Telephone-telephone condition

Figure 2: Comparison of EER values of pooled and concatenated total-variability approach based GPLDA and HTPLDA systems at different lengths of active speech for each enrolment and verification condition,(a) interview-interview, (b) interview-telephone, (c) telephone-interview *and* (d) telephone-telephone

Initially, experiments were conducted for telephone-only speaker verification, examining the performance of the GPLDA and HTPLDA systems compared with standard and truncated evaluation conditions. These experiments found that the HTPLDA system continued to achieved better performance than the GPLDA as the length of the truncated evaluation data decreased. The advantages of including short utterances in development were also investigated, finding that having short utterances available for normalisation and PLDA modeling provided an improvement in speaker verification performance when compared to development in full-length data. This approach is very useful in real world speaker verification applications because required development data can be reduced.

Finally, a small set of experiments were conducted to investigate the performance of mismatched and matched *interview-interview* enrolment and verification for both the GPLDA and HTPLDA approaches in short utterance evaluation conditions. These experiments compared two approaches to combing channel representations in the total-variability calculation, finding that improved performance can be obtained by pooling the development data, rather than concatenating two separately-trained total-variability spaces from each channel.

More recently it was found that HTPLDA technique can be replaced with length normalized GPLDA technique, since it is computationally less expensive and achieves similar performance as HTPLDA. In our future work, we will investigate length normalized GPLDA technique with limited data conditions.

## 6. Acknowledgements

This project was supported by the Cooperative Research Centre for Advanced Automotive Technologies (AutoCRC).

## 7. References

[1] R. Vogt, B. Baker, and S. Sridharan, "Factor analysis subspace estimation for speaker verification with short utterances," in *Interspeech 2008*, (Brisbane, Australia), pp. 853–856, September 2008.

[2] R. Vogt, C. Lustri, and S. Sridharan, "Factor analysis mod-

elling for speaker verification with short utterances," in *Odyssey: The Speaker and Language Recognition Workshop*, 2008.

[3] A. Kanagasundaram, R. Vogt, D. Dean, S. Sridharan, and M. Mason, "i-vector based speaker recognition on short utterances," in *Proceed. of INTERSPEECH*, pp. 2341–2344, International Speech Communication Association (ISCA), 2011.

[4] M. McLaren, R. Vogt, B. Baker, and S. Sridharan, "Experiments in SVM-based speaker verification using short utterances," in *Proc. Odyssey Workshop*, 2010.

[5] J. Price and T. Gee, "Face recognition using direct, weighted linear discriminant analysis and modular subspaces," *Pattern Recognition*, vol. 38, no. 2, pp. 209–219, 2005.

[6] P. Kenny, "Bayesian speaker verification with heavy tailed priors," in *Proc. Odyssey Speaker and Language Recogntion Workshop, Brno, Czech Republic*, 2010.

[7] M. Senoussaoui, P. Kenny, N. Brummer, E. de Villiers, and P. Dumouchel, "Mixture of PLDA models in i-vector space for gender independent speaker recognition," *Proceed. of INTERSPEECH*, pp. 25–28, 2011.

[8] L. Burget, O. Plchot, S. Cumani, O. Glembek, P. Matejka, and N. Brümmer, "Discriminatively trained probabilistic linear discriminant analysis for speaker verification," pp. 4832–4835, ICASSP, 2011.

[9] Z. Khan and F. Dellaert, "Robust generative subspace modeling: The subspace t distribution," 2004.

[10] N. Dehak, R. Dehak, P. Kenny, N. Brummer, P. Ouellet, and P. Dumouchel, "Support vector machines versus fast scoring in the low-dimensional total variability space for speaker verification," in *Proceedings of Interspeech*, p. 1559 1562, 2009.

[11] P. Kenny, P. Ouellet, N. Dehak, V. Gupta, and P. Dumouchel, "A study of inter-speaker variability in speaker verification," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 5, pp. 980–988, 2008.

[12] N. Dehak, R. Dehak, J. Glass, D. Reynolds, and P. Kenny, "Cosine similarity scoring without score normalization techniques," *Odyssey Speaker and Language Recognition Workshop*, 2010.

[13] M. McLaren and D. van Leeuwen, "Improved speaker recognition when using i-vectors from multiple speech sources," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pp. 5460–5463, 2011.

[14] S. Shum, N. Dehak, R. Dehak, and J. Glass, "Unsupervised speaker adaptation based on the cosine similarity for text-independent speaker verification," *Proc. Odyssey*, 2010.