# INTER DATASET VARIABILITY MODELING FOR SPEAKER RECOGNITION

*Hagai Aronowitz*

IBM Research - Haifa, Israel

## ABSTRACT

We introduce a novel approach of addressing inter-dataset variability in the context of speaker recognition in a mismatched condition under the JHU-2013 domain adaptation challenge (DAC) framework. Previously, we took a subspace removal approach for inter-dataset variability compensation (IDVC) of within speaker variability. In this work we substitute subspace removal with incorporation of the variability into the Probabilistic Linear Discriminant Analysis (PLDA) model. We do that by introducing a novel optimality criterion which is minimizing the expected square error in estimation of the log-likelihood ratio of target trials when dataset-dependent PLDA models are replaced by a dataset independent PLDA model. The result we obtain is a correction term for the commonly estimated within speaker variability matrix. The correction term represents the normalized inter-dataset variability of the within speaker variability matrices. The proposed method outperforms the extended IDVC method on the DAC.

***Index Terms***— robust speaker recognition, inter dataset variability compensation, domain adaptation challenge, inter dataset variability modeling,

## 1. INTRODUCTION

Current state-of-the-art systems in text independent speaker recognition obtain very low error rates [1] when trained on a large dataset with thousands of multi-session speakers as long as the development data matches the evaluation data. However, when development and evaluation data are mismatched, accuracy degrades dramatically [2].

The cross domain speaker recognition task was addressed in the JHU 2013 speaker recognition workshop [3] in the framework of the Domain Adaptation Challenges (DAC) [2]. The challenge was motivated by preliminary experiments that showed that a state-of-the-art i-vector PLDA system [4] built on the Switchboard [5] corpus had a 3 times larger equal error rate (EER) on the NIST 2010 SRE (condition 5) [6], compared to a system built on a subset of the MIXER corpus (NIST 2004-2008 SREs) [5] with a comparable size.

Recently, interest in cross domain speaker recognition increased significantly, and it is fundamental in the NIST 2016 speaker recognition evaluation [7].

Previous works addressing domain mismatch in speaker recognition may be categorized into three main approaches. The first approach is domain adaptation using some amount of adaptation data from the target domain, either labeled or unlabeled [8-14]. The second approach is improving domain robustness using the source domain data only (without the use of any target data) [15-19]. The third approach is retraining the system on a limited amount of labeled target data [10, 20-21].

In this paper we follow the second approach in which no target data is used, and therefore the focus is on improving the robustness of the system trained on the source data. We propose an improvement of our inter-dataset variability compensation (IDVC) method firstly introduced in [16] and extended in [17].

The basic IDVC method [16] aims at explicitly compensating dataset shift in the i-vector space as a pre-processing cleanup step, and was shown in [16, 11-12, 19] to be very effective on the DAC. Recently, it was also found effective by 7 participating sites for the NIST 2016 SRE [23-24]. The extended version introduced in [17] further improves performance on the DAC by compensating not only additive shifts in the i-vector space but also inter-dataset differences in within and between (across) speaker variability.

In this paper we withdraw the subspace removal technique for removing inter-dataset differences in within speaker variability. Instead, we optimize the PLDA model to account for inter-dataset variability. We present empirical results which indicate that our proposed method outperforms the extended IDVC method, and obtains the best published results on the DAC when no adaptation data is used. Furthermore, the results are competitive to other published works on the DAC even though they do use unlabeled adaptation data.

The rest of the paper is organized as follows: Section 2 provides an overview of the extended IDVC method. Section 3 describes our proposed method. Section 4 reports the experiments and results. Finally, Section 5 concludes.

## 2. EXTENDED INTER DATASET VARIABILITY COMPENSATION

IDVC aims at estimating and removing dataset mismatch in the i-vector domain. This is done by first partitioning the development data into subsets corresponding to different sources, and then training a PLDA model for each subset. The variability in the PLDA hyper-parameters across the subsets is analyzed and a low-dimensional subspace in the i-vector space accounting for most of that variability is pursued. The estimated low-subspace is then removed from all i-vectors as a pre-processing step before i-vector length normalization and whitening. The method is described in detail in the following subsections.

### 2.1. Two covariance model

The PLDA framework assumes that the i-vectors distribute according to Equation (1):

$$\phi = \mu + s + c \tag{1}$$

where $\phi$ denotes an i-vector, $s$ denotes a speaker component, $c$ denotes a channel (or within-speaker variability) component, and $\mu$ denotes the center of the i-vector space. Components $s$ and $c$ are

assumed to distribute normally with zero mean and covariance matrices B (between speaker) and W (within speaker) respectively.

The PLDA model is thus parameterized by $\{\mu, B, W\}$ and the goal of any PLDA training or adaptation algorithm is to estimate (or adapt) these hyper-parameters.

### 2.2. The extended IDVC method

We hypothesize that some directions in the i-vector space are more sensitive to dataset mismatch than other directions. In order to make a PLDA system robust to dataset mismatch, we aim at finding and removing a low-dimensional subspace (from the i-vector space) which is sensitive to dataset mismatch.

The mismatch-sensitive subspace can be estimated assuming that the development dataset is heterogeneous and we are able to partition it into a set of homogenous subsets. The homogenous subsets of the development dataset may be used to estimate a PLDA model for each subset independently, and the mismatch-sensitive subspace may be estimated from the collection of PLDA models.

Given a set of PLDA models parameterized by $\{\mu_i, B_i, W_i\}$ a subspace is estimated for each type of hyper-parameter independently. For $\mu$, the subspace is obtained by applying Principal Component Analysis (PCA) to the set of vectors $\{\mu_i\}$. For W and B, corresponding subspaces are obtained using the following procedure (shown for W):

1. Whiten the i-vector space with respect to the average within speaker variability covariance matrix $\overline{W} = \frac{1}{n}\sum_i W_i$
   - Calculate the square root of $\overline{W}^{-1}$ denoted by R
   - Define: $\hat{W}_i = RW_iR$
2. Compute $\Omega = \frac{1}{n}\sum_i \hat{W}_i^2$
3. Find the $k$ top eigenvectors of $\Omega$: $v_1,\ldots v_k$
4. Span the required subspace using $R^{-1}v_1,\ldots R^{-1}v_k$

More details can be found in [17].

## 3. INTER DATASET VARIABILITY MODELING (IDVM)

The main drawback we find in the extended IDVC scheme is that the subspace we remove may contain discriminative speaker information. Furthermore, the choice of the subspace rank has no theoretic basis and is tuned in practice on a tuning dataset.

For the $\mu$ hyper-parameter, experiments reported in [17] indicate that the problem is not very significant and setting the PCA subspace rank to the maximal possible rank ($n$-1 where $n$ is the number of subsets in the development dataset) seems to be optimal for the case of high dataset mismatch, and with a minimal degradation for the case of minimal dataset mismatch.

For the B hyper-parameter, we hypothesis that it is relatively insensitive to (DAC-like) dataset mismatch. We validate this hypothesis on the DAC data (Section 4).

The main contribution of this work is a novel modeling of the inter-dataset variability of the W hyper-parameter. Given subsets of the developments set with estimated within speaker variability covariance matrices $W_1,\ldots,W_n$, we aim at estimating a value (we denote by $\Lambda$) for the W hyper-parameter of the PLDA model which minimizes the expected inaccuracy of the estimated log-likelihood ratio (LLR) which is due to using $\Lambda$ instead of $W_i$. We describe in detail our approach in the following subsections.

### 3.1. Simplified PLDA scoring

Given i-vectors $x$ and $y$, we want to calculate the LLR with respect to the null hypothesis that $x$ and $y$ originate from different speakers.

PLDA gives a closed form expression for the LLR. Assuming that the i-vector space is centered ($\mu$=0), the LLR can be calculated as follows [4]:

$$llr = \log\frac{P(x,y|same)}{P(x,y|diff)} = x^T Qx + y^T Qy + 2x^T Py + \text{const} \quad (2)$$

with

$$\begin{aligned} Q &= T^{-1} - (T - BT^{-1}B)^{-1} \\ P &= T^{-1}B(T - BT^{-1}B)^{-1} \end{aligned} \quad (3)$$

where T is the total covariance matrix (T=B+W). As W is dataset dependent (but B is not according to our analysis), T, P and Q are also dataset dependent. We can therefore (for the sake of the analysis only) set B to be equal to the identity matrix I (by whitening the i-vector space with respect to B). We obtain:

$$\begin{aligned} Q &= T^{-1} - (T - T^{-1})^{-1} \\ P &= (T^2 - I)^{-1} \end{aligned} \quad (4)$$

For the sake of optimization of $\Lambda$ only (and not for actual scoring), we approximate Eq. (4) as follows:

$$\begin{aligned} Q &= (I + W)^{-1} - (I + W - (I + W)^{-1})^{-1} \approx I - W - \tfrac{1}{2}W^{-1} \approx I - \tfrac{1}{2}W^{-1} \\ P &= W^{-1}(2I + W)^{-1} \approx \tfrac{1}{2}W^{-1} \end{aligned} \quad (5)$$

Note that we obtain $Q \approx I - P$. The LLR (Eq. (2)) can be therefore approximated as:

$$llr \approx x^T x + y^T y - \tfrac{1}{2}(x\text{-}y)^T W^{-1}(x\text{-}y) + \text{const}. \quad (6)$$

Let $\delta$=$x$-$y$. For dataset $i$, we can estimate the LLR either using a dataset dependent (DD) matrix $W_i$ (we denote this LLR with $llr_{DD}$) or using a dataset independent (DI) matrix $\Lambda$ (we denote this LLR with $llr_{DI}$):

$$llr_{DD}(x,y,i) = x^T x + y^T y - \tfrac{1}{2}\delta^t W_i^{-1}\delta + \text{const} \quad (7)$$

$$llr_{DI}(x,y,i) = x^T x + y^T y - \tfrac{1}{2}\delta^t \Lambda^{-1}\delta + \text{const} \quad (8)$$

The difference between DD-LLR and DI-LLR is the LLR estimation error. We demote it by $e(\delta, i, \Lambda)$:

$$e(\delta, i, \Lambda) = \delta^t (W_i^{-1} - \Lambda^{-1})\delta. \quad (9)$$

Finally, the loss function L is defines as the square of the LLR estimation error:

$$\mathrm{L}(\delta,i,\Lambda)=e^{2}(\delta,i,\Lambda) \quad (10)$$

We are interested in minimizing the mean over different datasets of the expected loss of target trials:

$$\hat{\Lambda}=\arg\min_{\Lambda}\tfrac{1}{n}\sum_{i}\mathrm{E}_{\delta}\left[\mathrm{L}(\delta,i,\Lambda)\right] \quad (11)$$

We show in the Appendix that the solution to Eq. (11) is:

$$\Lambda=\left(\tfrac{1}{n}\sum_{i}\mathrm{W}_{i}^{2}\right)\left(\tfrac{1}{n}\sum_{i}\mathrm{W}_{i}\right)^{-1}=\overline{\mathrm{W}}+\left(\tfrac{1}{n}\sum_{i}\left(\mathrm{W}_{\mathrm{i}}-\overline{\mathrm{W}}\right)^{2}\right)\overline{\mathrm{W}}^{-1} \quad (12)$$

The optimal $\Lambda$ is therefore equal to the average within speaker covariance matrix plus a term that represents the variability of the within speaker covariance matrix across different datasets, normalized by the average within speaker covariance matrix.

## 4. EXPERIMENTS AND RESULTS

### 4.1. Datasets

We use the JHU-2013 speaker recognition workshop DAC setup which can be downloaded from [3]. The source development dataset (Switchboard) consists of all telephone calls taken from Switchboard-I and Switchboard-II. The dataset consists of 3114 speakers and 33039 sessions. The target development dataset (MIXER) consists of a subset of telephone calls taken from SREs 2004-2008. The dataset consists of 3790 speakers and 36470 sessions. The NIST 2010 SRE condition 5 core extended trial list is used for evaluation. The dataset consists of 7169 target trials and 408956 impostor trials. We report results by pooling male and female trials. Three error measures are used: EER, DCF (08') and DCF (10') as specified in [6].

For the use of the IDVC and the IDVM methods, Switchboard is partitioned into 12 gender dependent subsets, based on 6 different releases (2 cellular releases, and 4 landline). More details can be found in [16].

### 4.2. Baseline system

The i-vectors we use are those supplied by the DAC organizers and may be downloaded from [2]. A detailed description of their creation can be found in [10]. We center the i-vectors using the Switchboard development data. Prior to PLDA modeling, the dimensionality of the i-vectors is reduced using GI-LDA to 400. The next steps are within class covariance normalization (WCCN) [3] and length normalization [3]. Gender-dependent (GD) PLDA is then applied with full rank between and within covariance matrices. For the sake of flexibility in our experiments, we do not use EM to estimate the PLDA model. Instead we use Maximum Likelihood (ML) followed by a cleanup step with removes expected estimation biases and smoothing of the off-diagonal covariance elements with the diagonal elements. We observed a slight improvement in the baseline due to this approach compared to the standard EM training we used in [17]. Finally, we compute a single bias term (using Switchboard) for the female scores in order to calibrate them to the male scores.

### 4.3. Between speaker variability analysis

Table 1. *Results for pooled male and female trials without IDVC.*

| Devset | EER [%] | DCF-08 | DCF-10 |
|---|---|---|---|
| **Switchboard** | 8.16 | 0.323 | 0.683 |
| **Mixer** | 2.23 | 0.111 | 0.378 |
| **Mixer (Switchboard for B)** | 2.28 | 0.118 | 0.377 |

Table 2. *Results for pooled male and female trials with IDVC compared to the proposed method*

| Configuration | EER [%] | DCF-08 | DCF-10 |
|---|---|---|---|
| **IDVC: $\mu$ only** | 3.68 | 0.184 | 0.531 |
| **IDVC: $\mu$ , W (rank=100)** | 3.03 | 0.146 | 0.477 |
| **IDVM** | 2.82 | 0.141 | 0.467 |

Table 3. *Results comparing the proposed method (IDVM) to related published works, all using the same i-vectors provided within the DAC framework. The first 5 systems do not use unlabeled target data for training, while the last 3 system do.*

| System | Target data used? | EER [%] Male+ Female | EER [%] Male | EER [%] Female |
|---|---|---|---|---|
| **IDVM** | | 2.82 | 2.54 | 3.10 |
| **IDVC: $\mu$ only [16]** | | 3.68 | 3.32 | 4.12 |
| **IDVC: $\mu$ + W100 [17]** | No | 3.03 | 2.86 | 3.22 |
| **WCC [18]** | | 5.04 | - | - |
| **Library of whiteners [19]** | | 3.86 | - | - |
| **WCC-8G [18]** | | 2.91 | - | - |
| **Infomap+AHC [9]** | Yes | 2.53 | | |
| **IB-clustering [12]** | | - | 2.5 | - |

In order to assess the sensitivity of the B matrix to dataset variability we ran an experiment where the $\mu$ and W PLDA hyper-parameters are estimated from the target devset (MIXER) and only B is estimated from the source devset (Switchboard). The results are compared in Table 1 to the baseline system trained on either Switchboard or MIXER. Results indicate that B is not very sensitive to dataset mismatch (~2% relative degradation).

### 4.4. IDVM analysis

The proposed IDVM method is compared to basic IDVC and to extended IDVC. Note that running IDVM includes running basic IDVC for compensating inter-dataset variability in the $\mu$ hyper-parameter. The results are reported in Table 2. Results indicate that IDVM obtains best results. EER reduction is 24% and 7% compared to basic and extended IDVC respectively. The DCF reductions are smaller. Moreover, contrary to extended IDVC which requires tuning the rank of the compensated subspace (see [17]), IDVM does not require tuning.

Table 3 reports a comparison of the IDVM method to other methods reported in the literature in the framework of the DAC (methods that do not use the standard i-vectors set are excluded from the comparison). IDVM outperforms all other reported methods that do not use any target data for training, and obtains comparable results to [8] and [12] which do use unlabeled target development data (while IDVM does not use any target data whatsoever). Finally, the unsupervised clustering method of unlabeled target data reported in [9] outperforms IDVM by 10% relative (but uses unlabeled target data).

## 5. CONCLUSIONS

In this work we investigate how to optimally apply PLDA-based speaker recognition when the development data is heterogeneous and mismatches the target domain in the sense that the within speaker variability varies significantly. We observe that on the JHU 13' Domain Adaptation Challenge, this is a major source of degradation on top of the dataset shift (centering) issue we address in the basic IDVC method.

As opposed to the subspace removal approach taken by the extended IDVC method, we address the inter-dataset variability of within speaker variability by finding an optimal value for the within speaker variability covariance matrix W in the PLDA model. We define a novel optimality criterion which is minimizing the expected square error in estimation of the log-likelihood ratio of target trials when dataset-dependent PLDA models are replaced by a dataset independent PLDA model.

It turns out that using some reasonable approximations, there is a simple closed form expression for the optimal W which is taking the average of the dataset-dependent W values and adding a correction term which represents the normalized inter-dataset variability of W. For a homogeneous development set our solution converges back to the commonly used average.

The proposed method named inter dataset variability modeling (IDVM) was found to be effective on the DAC, and it outperforms significantly other methods that follow the same protocol (namely using the DAC i-vectors and not training on the unlabeled MIXER).

## 6. APPENDIX

We rewrite the difference between DD-LLR and DI-LLR (Eq. (9)) as following:

$$\mathrm{e}(\xi,i,\Lambda)=\xi^t\left(\mathrm{I}-\mathrm{W}_i^{\frac{1}{2}}\Lambda^{-1}\mathrm{W}_i^{\frac{1}{2}}\right)\xi \tag{13}$$

where $\xi$ is a random vector in the i-vector space with a multivariate standard normal distribution. Matrix $\mathrm{I}-\mathrm{W}_i^{\frac{1}{2}}\Lambda^{-1}\mathrm{W}_i^{\frac{1}{2}}$ can be factored as

$$\mathrm{I}-\mathrm{W}_i^{\frac{1}{2}}\Lambda^{-1}\mathrm{W}_i^{\frac{1}{2}}=\mathrm{P}^t\Omega\mathrm{P} \tag{14}$$

where P is an orthogonal matrix and $\Omega$ is diagonal with diagonal elements $\lambda_1,\ldots,\lambda_d$. Equation (13) can be reformulated as:

$$\mathrm{e}(\xi,i,\Lambda)=(\mathrm{P}\xi)^t\Omega(\mathrm{P}\xi) \tag{15}$$

and the expected loss as:

$$\begin{aligned}\mathrm{E}_\xi\left[\mathrm{e}(\delta,i,\Lambda)^2\right]&=\mathrm{E}_\xi\left[(\mathrm{P}\xi)^t\Omega(\mathrm{P}\xi)(\mathrm{P}\xi)^t\Omega(\mathrm{P}\xi)\right]\\&=c\sum_j\lambda_j^2\\&=c\cdot\mathrm{tr}\left(\left(\mathrm{I}-\mathrm{W}_i^{\frac{1}{2}}\Lambda^{-1}\mathrm{W}_i^{\frac{1}{2}}\right)^2\right)\end{aligned}. \tag{16}$$

where $c$ is a constant taken from the chi-distribution. Due to the additivity of the trace function and the trace permutation rule for a product of symmetric matrices, we get

$$\begin{aligned}&\mathrm{tr}\left(\left(\mathrm{I}-\mathrm{W}_i^{\frac{1}{2}}\Lambda^{-1}\mathrm{W}_i^{\frac{1}{2}}\right)^2\right)=\\&\mathrm{tr}\left(\mathrm{I}+\mathrm{W}_i^{\frac{1}{2}}\Lambda^{-1}\mathrm{W}_i\Lambda^{-1}\mathrm{W}_i^{\frac{1}{2}}-2\mathrm{W}_i^{\frac{1}{2}}\Lambda^{-1}\mathrm{W}_i^{\frac{1}{2}}\right)=\\&\mathrm{tr}(\mathrm{I})+\mathrm{tr}\left(\mathrm{W}_i^2\Lambda^{-2}\right)-2\mathrm{tr}\left(\mathrm{W}_i\Lambda^{-1}\right)\end{aligned} \tag{17}$$

In order to minimize the expected loss L we take the derivative of the expected loss with respect to $\Lambda$, and equate it to zero:

$$\frac{\partial}{\partial\Lambda^{-1}}\mathrm{L}(\delta,i,\Lambda)=\frac{\partial}{\partial\Lambda^{-1}}\mathrm{tr}\left(\left(\mathrm{I}-\mathrm{W}_i^{\frac{1}{2}}\Lambda^{-1}\mathrm{W}_i^{\frac{1}{2}}\right)^2\right)=0 \tag{18}$$

Note that

$$\left(\mathrm{I}-\mathrm{W}_i^{\frac{1}{2}}\Lambda^{-1}\mathrm{W}_i^{\frac{1}{2}}\right)^2=\mathrm{I}-2\mathrm{W}_i^{\frac{1}{2}}\Lambda^{-1}\mathrm{W}_i^{\frac{1}{2}}+\mathrm{W}_i^{\frac{1}{2}}\Lambda^{-1}\mathrm{W}_i\Lambda^{-1}\mathrm{W}_i^{\frac{1}{2}} \tag{19}$$

and due to the symmetry of $\mathrm{W}_i$ and $\Lambda$:

$$\mathrm{tr}\left(\mathrm{I}-\mathrm{W}_i^{\frac{1}{2}}\Lambda^{-1}\mathrm{W}_i^{\frac{1}{2}}\right)^2=\mathrm{tr}(\mathrm{I})-2\mathrm{tr}\left(\mathrm{W}_i\Lambda^{-1}\right)+\mathrm{tr}\left(\mathrm{W}_i^2\Lambda^{-2}\right). \tag{20}$$

According to subsection 2.5 in [22]:

$$\frac{\partial}{\partial\Lambda^{-1}}\left(\mathrm{tr}(\mathrm{I})-2\mathrm{tr}\left(\mathrm{W}_i\Lambda^{-1}\right)+\mathrm{tr}\left(\mathrm{W}_i^2\Lambda^{-2}\right)\right)=2\Lambda^{-1}\mathrm{W}_i^2-2\mathrm{W}_i \tag{21}$$

Therefore,

$$\frac{1}{n}\sum_i\left(2\Lambda^{-1}\mathrm{W}_i^2-2\mathrm{W}_i\right)=0 \tag{22}$$

And the solution for $\Lambda$ is

$$\Lambda=\left(\frac{1}{n}\sum_i\mathrm{W}_i^2\right)\left(\frac{1}{n}\sum_i\mathrm{W}_i\right)^{-1}=\overline{\mathrm{W}}+\left(\frac{1}{n}\sum_i\left(\mathrm{W}_i-\overline{\mathrm{W}}\right)^2\right)\overline{\mathrm{W}}^{-1} \tag{23}$$

## 7. REFERENCES


[1] S.O. Sadjadi, S. Ganapathy and J. Pelecanos, "The IBM 2016 Speaker Recognition System", in Proc. *Speaker Odyssey*, 2016.

[2] Domain Adaptation Challenge 2013, Available online: http://www.clsp.jhu.edu/wp-content/uploads/sites/75/2015/10/WS13-Speaker-DAC.pdf, 2013.

[3] JHU 2013 speaker recognition workshop. Available online: http://www.clsp.jhu.edu/workshops/archive/ws13-summer-workshop/groups/spk-13/.

[4] D. Garcia-Romero and C. Y. Espy-Wilson, "Analysis of i-vector length normalization in speaker recognition systems," in Proc. *Interspeech*. 2011.

[5] The Linguistic Data Consortium (LDC) catalog. Available online: http://catalog.ldc.upenn.edu/project_index.jsp

[6] NIST 2010 SRE evaluation plan. Available online: http://www.nist.gov/itl/iad/mig/upload/NIST_SRE10_evalplan-r6.pdf.

[7] NIST 2016 SRE evaluation plan. Available online: https://www.nist.gov/sites/default/files/documents/itl/iad/mig/SRE16_Eval_Plan_V1-0.pdf

[8] M. Senoussaoui, P. Kenny, N. Dehak, and P. Dumouchel, "An i-vector Extractor Suitable for Speaker Recognition with both Microphone and Telephone Speech", in Proc *Speaker Odyssey*, 2010.

[9] D. Garcia-Romero, A. McCree, S. Shum, N. Brummer, and C. Vaquero, "Unsupervised Domain Adaptation for i-vector Speaker Recognition," in Proceedings of Speaker Odyssey,2014.

[10] D. Garcia-Romero and A. McCree, "Supervised domain adaptation for i-Vector based speaker recognition", in Proc. *ICASSP*, 2014.

[11] D. Garcia-Romero, X. Zhang, A. McCree, and D. Povey, "Improving Speaker Recognition Performance in the Domain Adaptation Challenge using Deep Neural Networks", in Proc. of *SLT*, 2014.

[12] S. Dey, S. Madikeri and P. Motlicek,"Information theoretic clustering for unsupervised domain adaptation", in Proc. of *ICASSP*, 2016.

[13] H. Aronowitz, O. Barkan, "On Leveraging Conversational Data for Building a Text Dependent Speaker Verification System", in Proc. *Interspeech*, 2013.

[14] H. Aronowitz, "Text Dependent Speaker Verification Using a Small Development Set", in Proc. *Speaker Odyssey*, 2012.

[15] M. McLaren and D. van Leeuwen, "Source-normalized LDA for robust speaker recognition using i-vectors from multiple speech sources", IEEE *Trans. Audio, Speech and Language Processing*, 20(3):755–766, March 2012.

[16] H. Aronowitz, "Inter dataset Variability compensation for speaker recognition", in Proc. *ICASSP*, 2014.

[17] H. Aronowitz "Compensating Inter-Dataset Variability in PLDA Hyper-Parameters for Robust Speaker Recognition", in Proc. *Speaker Odyssey*, 2014

[18] O. Glembek, J. Ma, P. Matejka, B. Zhang, O. Plchot, L. Burget, and S. Matsoukas, "Domain adaptation via within-class covariance correction in i-vector based speaker recognition systems," in Proc. *ICASSP*, 2014.

[19] E. Singer, D. Reynolds, "Domain mismatch compensation for speaker recognition using a library of whiteners," IEEE Signal Processing Letters, vol. 22, no. 11, pp. 2000–2003, 2015.

[20] H. Aronowitz, "Exploiting Supervector Structure for Speaker Recognition Trained on a Small Development Set", in Proc. *Interspeech*, 2015.

[21] H. Aronowitz, "Speaker Recognition using Matched Filters", in Proc. *ICASSP*, 2016.

[22] K.B. Petersen, M.S. Pedersen, The Matrix Cookbook, 2012. Available online: https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf.

[23] K.A. Lee et al., "The I4U submission to the 2016 NIST speaker recognition evaluation", 2016. Available online: http://waadbenkheder.fr/wp-content/uploads/2015/05/SRE_2016_I4U.pdf

[24] M. Rouvier, P.-M. Bousquet, M. Ajili, W. B. Kheder, D. Matrouf, J.-F. Bonastre, "LIA system description for NIST SRE 2016", 2016. Available online: https://arxiv.org/pdf/1612.05168.pdf**.**