*INTERSPEECH 2016*
September 8–12, 2016, San Francisco, USA

# LIA system for the SITW Speaker Recognition Challenge

*Waad Ben Kheder, Moez Ajili, Pierre-Michel Bousquet, Driss Matrouf and Jean-François Bonastre*

LIA, University of Avignon, France

## Abstract

This paper presents the speaker verification systems developed in the LIA lab at the University of Avignon for the SITW (Speakers In The Wild) challenge. We present the algorithms used to deal with additive noise, short utterances and propose an improved scoring scheme using a discriminative classifier and integrating the homogeneity of the two compared recordings. Due to the heterogeneity of this database (presence of background noise, reverberation, Lombard effect, etc.), it is hard to analyze the contribution of individual techniques used to deal with each problem. For this reason, a subset of the trials will be studied for each algorithm in order to emphasize its contribution.

**Index Terms**: speaker recognition, i-vector, sitw.

## 1. Introduction

Inspired by the joint factor analysis model proposed by Kenny [1], the i-vector framework has become a standard in speaker recognition (SR) systems [2–4] and multiple techniques have been developed to complement it in order to deal with both environment disturbances (noise, echo,..) and undesired variability sources (intra-speaker and channel variability).

The introduction of several normalization techniques [5–8] and the development a Gaussian backend termed PLDA [9, 10] (Probabilistic linear discriminant analysis) allowed SR systems to account for speaker and session variability in the i-vector space and achieve high recognition rates in clean conditions.

Dealing with noise has also been one of the principal areas of interest and different techniques have been proposed to deal with it in different domains. In the temporal domain, speech enhancement techniques [11–13] have been proven to be noise and SNR-level dependent yielding low improvement rates (10% of relative EER improvement). In the cepstral domain, several stochastic compensation algorithms have been proposed in [14] and have been shown to be effective when prior knowledge about test noise is available. Other techniques aim at building robust i-vector extractors using VTS-based algorithms [15]. In these algorithms, a non-linear noise model is used in the cepstral domain to model the relationship between clean and noisy cepstral coefficients. In the recognition phase, the developed noise model is integrated in the i-vector extractor to help estimate a ”cleaned-up” version of noisy i-vectors. Other algorithms use uncertainty propagation [16, 17] in order to make the i-vector extraction system focus on reliable or reliably enhanced features but such techniques showed little improvement compared to a baseline system performance. In the scoring phase, robust versions of the PLDA model have been proposed such as the ”multi-style” training regime [18] and the SNR-invariant version of the PLDA [19] where i-vectors extracted from utterances falling within a narrow SNR range are assumed to share similar SNR-specific information. Such systems achieve a relative improvement of 25% in EER compared to regular PLDA.

With the rise of deep learning in the last few years, DNN-based techniques have also been widely used in this context to either enhance cepstral features [20] or compute better i-vector statistics [21, 22]. This class of techniques has been shown to improve the recognition performance by up to 30%.

The recently published SITW (Speakers in the wild) database [23] provides an interesting dataset for the analysis and testing of new algorithms. In this paper, we explore two axes. On one side, we aim at improving the system performance by using two different techniques: 1 - *The I-MAP algorithm* [24–26] which is an i-vector denoising procedure based on an additive noise model in the i-vector space. It uses a Gaussian modeling of both clean i-vectors and the noise distributions in the i-vector space and have been proven to yield up to 60% of relative EER improvement compared to a baseline system performance. 2 - *Discriminative classifier specific to normalized i-vectors* which is an algorithm intended to improve PLDA metaparameters. This classifier, referred to as Orthonormal Discriminative classifier, is a novelty in the field. It extracts discriminative axes by using iterations of the Fisher criterion and improves the performance of the PLDA model by preventing the over-fitting problem with regard to the development data.

On the other side, we study the reliability of the SR system output. This part focuses on exploring the homogeneity of information between the two sides of a voice comparison trial. The homogeneity measure used in this paper is *NHM*, the information theory-based criterion we proposed recently [27, 28]. The impact of homogeneity on reliability has been studied using FABIOLE [29] and NIST [30] which motivates us to validate its use on SITW.

This paper is structured as follows; Section 2 presents the i-vector framework and the PLDA scoring model. Section 3 presents the I-MAP denoising technique. Section 4 presents the discriminative classifier used to improve PLDA hyper-parameters. Section 5 defines the homogeneity measure in speaker recognition context and Section 6 presents the experiments and results relative to each algorithm.

## 2. The i-vector paradigm

An i-vector extractor converts a sequence of acoustic vectors into a single low-dimensional vector representing the whole speech utterance. The speaker- and session-dependent super-vector $s$ of concatenated Gaussian Mixture Model (GMM) means is assumed to obey a linear model of the form:

$$s = m + Tw \tag{1}$$

where $m$ is the mean super-vector of the Universal Background Model (UBM), $T$ is the low-rank variability matrix obtained from a large dataset by MAP estimation [1] and $w$ is a normally distributed latent variable called ”i-vector”.

Extracting an i-vector from the total variability subspace is essentially a maximum a-posteriori adaptation of $w$ in the space

http://dx.doi.org/10.21437/Interspeech.2016-1310

defined by $T$ . The algorithms for the estimation of $T$ and the extraction of i-vectors are described in [31].

### 2.1. The PLDA model for i-vector scoring

In the i-vector framework, the problem of intersession variability is deferred to the scoring stage. The Gaussian Probabilistic Linear Discriminant Analysis (PLDA) was introduced in [5] as a generative i-vector model which assumes that each $d$-dimensional i-vector $w$ of a speaker $s$ can be decomposed as:

$$\mathbf{w} = \mu + \mathbf{\Phi y}_s + \varepsilon \qquad (2)$$

where $\mathbf{\Phi y}_s$ and $\varepsilon$ are assumed to be statistically independent and $\varepsilon$ follows a centered Gaussian distribution with full covariance matrix $\mathbf{\Lambda}$. Speaker factor $\mathbf{y}_s$ can be a full-rank $d$-vector (this model is referred to as *two-covariance model* [32]) or constrained to lie in the $r$-linear range of the $d \times r$ matrix $\mathbf{\Phi}$, referred to as *eigenvoice subspace* [33].

After estimation of the PLDA meta-parameters, the speaker verification score given two i-vectors $w_1$ and $w_2$ is the likelihood-ratio described by Equation 3, where the hypothesis $\theta_{tar}$ states that inputs $w_1$ and $w_2$ are from the same speaker and the hypothesis $\theta_{non}$ states they are from different speakers.

$$score = log \frac{P(w_1, w_2|\theta_{tar})}{P(w_1, w_2|\theta_{non})} \qquad (3)$$

## 3. The I-MAP denoising procedure

In our previous work [24–26], we proposed an additive noise model in the i-vector space represented by the equation:

$$N = Y - X \qquad (4)$$

Where $X$ and $Y$ are two random variables representing respectively clean and noisy i-vectors and $N$ represents the noise. Using full-covariance Gaussian distributions for both clean i-vectors $d_X \sim \mathcal{N}(X; \mu_X, \Sigma_X)$ and noise in the i-vector space $d_N \sim \mathcal{N}(N; \mu_N, \Sigma_N)$, it is possible to write the cleaned-up version $\hat{X_0}$ of a noisy i-vector $Y_0$ using MAP criterion as [24–26]:

$$\hat{X_0} = (\Sigma_N^{-1} + \Sigma_X^{-1})^{-1}(\Sigma_N^{-1}(Y_0 - \mu_N) + \Sigma_X^{-1}\mu_X) \qquad (5)$$

The derivation of Equation 5 is detailed in [24–26].

### 3.1. Estimation of $\mathcal{N}(X; \mu_X, \Sigma_X)$ and $\mathcal{N}(N; \mu_N, \Sigma_N)$

As detailed in [24, 25], the clean i-vectors distribution $\mathcal{N}(X; \mu_X, \Sigma_X)$ and the noise distribution $\mathcal{N}(N; \mu_N, \Sigma_N)$ are two key components in the I-MAP procedure.

Since $\mathcal{N}(X; \mu_X, \Sigma_X)$ is noise-independent, it can be estimated once and for all over a large set of clean i-vectors in an off-line step initially before performing any compensation.

On the other hand, $\mathcal{N}(N; \mu_N, \Sigma_N)$ makes the system able to adapt to the noise present in the signal and compensate its effect more effectively. It is estimated for each different test noise and it requires the existence of clean i-vectors and the noisy versions corresponding to the same segments. First, for the clean part and once the train set is fixed, the corresponding clean i-vectors $(X)$ are extracted. Then, for a given noisy test segment, the noise is extracted from the signal (using a VAD system and selecting the low-energy frames) then added to the clean train set in the time domain. Finally, the corresponding noisy i-vectors $(Y)$ are estimated and Equation (5) is used to compute $N$ then $\mathcal{N}(N; \mu_N, \Sigma_N)$. The full algorithm is shown in Figure 1 and more details about the technique can be found in [24].

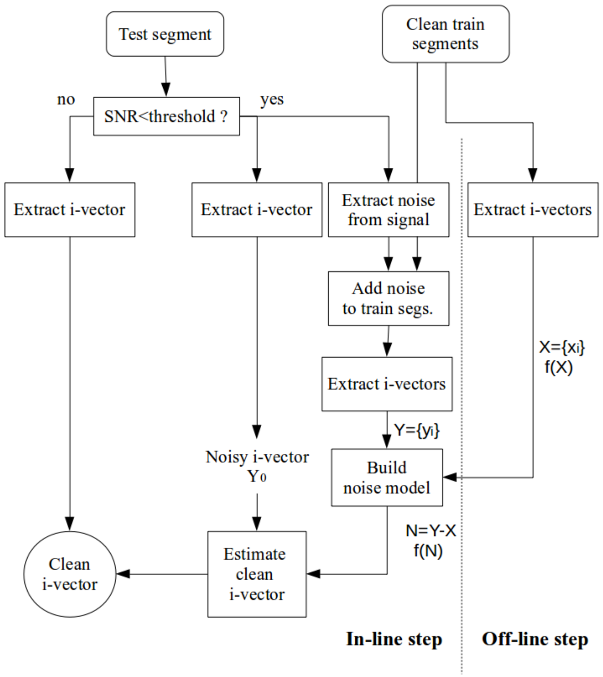

Figure 1: I-vector cleaning procedure using I-MAP.

## 4. Discriminative classifier specific to normalized i-vectors

LIA system 2 for SITW is based on a new discriminative classifier for speaker recognition, specific to normalized i-vectors. This method seeks to optimize PLDA matricial parameters. Unlike usual discriminative approaches for speaker recognition, which rely on minimization of a cross-entropy function by using logistic regression, this method uses a specific algorithm based on Fisher criterion. To overcome the well known issue of over-fitting to development data, a new procedure is added to the pre-normalization step, which limits the amount of coefficient to discriminatively train. Detailed description of this method can be found in [34].

### 4.1. Additional normalization procedure

Once i-vectors have been normalized, their length is equal to 1, but it is also worth noting that their within-class covariance matrix $\mathbf{W}$ is almost exactly isotropic [6], i.e. $\mathbf{W} \approx \sigma\mathbf{I}$ where $\sigma$ is a positive real and $\mathbf{I}$ is the identity matrix. We propose to add a supplementary step to the normalization procedure. I-vectors of training and test are rotated by the eigenvector basis of between-class covariance matrix of the training dataset. By this way, the new between-class covariance matrix is diagonal, with its diagonal equal to its eigenvalue spectrum. The new within-class covariance matrix remains almost isotropic (and therefore diagonal), as the eigenvector basis is orthogonal. Furthermore, it is shown in [34] that, after this additional procedure, matrices $\mathbf{\Phi\Phi}^t$ and $\mathbf{\Lambda}$ of PLDA modeling become almost diagonal, and even isotropic for $\mathbf{\Lambda}$. Given two i-vectors $\mathbf{w}_i$, $\mathbf{w}_j$, the PLDA-based log-likelihood ratio can be written as:

$$
s_{i,j} = \sum_{k=1}^{r} \left\{ \begin{array}{c} p_k \mathbf{w}_{i,k} \mathbf{w}_{j,k} + \frac{1}{2} q_k \left( \mathbf{w}_{i,k}^2 + \mathbf{w}_{j,k}^2 \right) \\ - \left( p_k + q_k \right) \mu_k \left( \mathbf{w}_{i,k} + \mathbf{w}_{j,k} \right) \end{array} \right\} \\ + res_{i,j} \tag{6}
$$

where $r$ is the rank of the PLDA eigenvoice subspace, $p, q \in \mathbb{R}^d$ and the residual term $res_{i,j}$ sums all the diagonal terms beyond the $r^{th}$ dimension, all the *off-diagonal* terms and offsets. It is also shown in [34] that, after the additional procedure, the residual term $res_{i,j}$ is negligible in regard to the first $r$ diagonal terms of the score. By this way, a discriminative classifier intended to optimize PLDA parameters, which is initially of order the square of the i-vector size, can be replaced by a constrained discriminative classifier of low order with a minimal loss of accuracy.

### 4.2. Orthonormal discriminative classifier

Let us define the expanded $\mathbb{R}^{r+1}$ vector $\varphi_{i,j}$ of a trial $(\mathbf{w}_i, \mathbf{w}_j)$ by:

$$
\varphi_{i,j} = \left[ \begin{array}{c} \left\{ \begin{array}{c} p^{(r)} \circ \mathbf{w}_i^{(r)} \circ \mathbf{w}_j^{(r)} \\ +\frac{1}{2} q^{(r)} \circ \left( \mathbf{w}_i^{(r)} \circ \mathbf{w}_i^{(r)} + \mathbf{w}_j^{(r)} \circ \mathbf{w}_j^{(r)} \right) \\ -\mu^{(r)} \circ \left( p^{(r)} + q^{(r)} \right) \circ \left( \mathbf{w}_i^{(r)} + \mathbf{w}_j^{(r)} \right) \end{array} \right\} \\ , \\ res_{i,j} \end{array} \right] \tag{7}
$$

where the superscript $^{(r)}$ indicates the first $r$ components of a vector and the symbol $\circ$ denotes the element wise product. Score of (6) can be written as:

$$
s_{i,j} = \varphi_{i,j}^t . \mathbf{1}_{r+1} \tag{8}
$$

where $\mathbf{1}_{r+1}$ is the $\mathbb{R}^{r+1}$ vector of ones. As formulated, PLDA score is, geometrically, the projection of an expanded vector $\varphi_{i,j}$ onto the axis $\mathbf{1}_{r+1}$. Unlike the usual discriminative classifiers, which attempt to find out a unique normal vector of a separation hyperplane, we propose to extract a discriminant subspace (by decreasing variance, in a way similar to singular value decomposition), then to combine its basis to find out the unique normal vector needed by speaker detection.

Introduced in [35], and successfully applied in fields such as face recognition [36,37], "Orthonormal Discriminative (OD) classifier" is a discriminative method based on Fisher criterion, which allows to extract more axes than classes. Given a training corpus $\mathcal{T}$ of target and non-target trial expanded vectors, the following algorithm describes this method:

```
for $k = 1$ to $K$
    Compute target and non-target means $g_t^{(k)}, g_n^{(k)}$ of $\mathcal{T}$
    Compute covariance matrices $\mathcal{W}_t^{(k)}, \mathcal{W}_n^{(k)}$ of $\mathcal{T}$
    Compute between-class covariance matrix $\mathcal{B}$ of $\mathcal{T}$
    Extract vector $u^{(k)}$ maximizing the
    Fisher criterion $\frac{v^t \mathcal{B} v}{v^t \mathcal{W} v}$
    Project $\mathcal{T}$ onto the orthogonal subspace of $u^{(k)}$.
```

The final normal vector, which replaces the vector $\mathbf{1}_{r+1}$ in (8), is the following weighted sum of extracted vectors:

$$
u = \sum_{k=1}^{K} \left( \alpha_t \mathcal{W}_t^{(k)} + \alpha_n \mathcal{W}_n^{(k)} \right)^{-1} \left( g_t^{(k)} - g_n^{(k)} \right) \tag{9}
$$

where $\alpha_t, \alpha_n$ denote the target and non-target priors of $\mathcal{T}$, respectively.

As far as training complexity is concerned, OD training also has the advantage of being very thrifty in terms of time and memory requirements [34].

## 5. Homogeneity measure

In this section, we introduce $HM()$, the *information theory* (IT) based acoustic homogeneity measure we presented in [27]. Its objective is to calculate the amount of acoustic information that appertains to the same (acoustic) class between the two voice records. The set of acoustic frames gathered from the two files $S_A$ and $S_B$ is decomposed into acoustic classes thanks to a Gaussian mixture Model (GMM) clustering. Then the homogeneity is first estimated in term of bits as the amount of information embed by the respective "number of acoustic frames" of $S_A$ and $S_B$ linked to a given acoustic class. Each acoustic class is represented by the corresponding Gaussian component of the GMM model. The occupation vector could be seen as the number of acoustic frames of a given recording belonging to each class $m$. It is noted: $[\gamma_{g_m}(s)]_{m=1}^M$.

Given a Gaussian $g_m$ and two posterior probability vectors of the two voice records $S_A$ and $S_B$, $[\gamma_{g_m}(A)]_{m=1}^M$ and $[\gamma_{g_m}(B)]_{m=1}^M$, we define also:

- $\chi_A \cup \chi_B = \{x_{1A}, ...., x_{NA}\} \cup \{x_{1B}, ...., x_{NB}\}$ the full data set of $S_A$ and $S_B$ with cardinality N=$N_A$ +$N_B$.
- $\gamma(m)$ and $\omega(m)$ are respectively the occupation and the prior of Gaussian $m$ where $\omega(m) = \frac{\gamma(m)}{\sum_{k=1}^M \gamma(k)} = \frac{\gamma(m)}{N}$.
- $\gamma_A(m)$ (respectively $\gamma_B(m)$ ) is the partial occupations of the $m^{th}$ component due to the voice records $S_A$ (respectively $S_B$).
- $p_m$ is the probability of the Bernoulli distribution of the $m^{th}$ bit (due to the $m^{th}$ component), B$(p_m)$. $p_m$=$\frac{\gamma_A(m)}{\gamma(m)}$ , $\overline{p}_m = 1 - p_m = \frac{\gamma_B(m)}{\gamma(m)}$.
- $H(p_m)$ the entropy of the $m^{th}$ Gaussian (the unit is bits) given by: $H(p_m) = -p_m log_2(p_m) - \overline{p}_m log_2(\overline{p}_m)$.

### 5.1. Non-normalized Homogeneity Measure (NHM())

In this paper, we use the *Non-normalized Homogeneity Measure* ($NHM()$) proposed in [27, 28] which calculates the quantity of homogeneous information between the two voice records as shown in Equation 10. The amount of information is defined in term of number of acoustic frames. $NHM()$ measures the *Bic Entropy Expectation* BEE with respect of the quantity of information present in each acoustic class $\{\gamma(m)\}_{i=1}^M$.

$$
\begin{aligned} NHM_{BEE} &= \sum_{m=1}^{M} (\gamma_A(m) + \gamma_B(m)) H(p_m) \\ &= \sum_{m=1}^{M} \gamma(m) H(p_m) \end{aligned} \tag{10}
$$

## 6. Experiments and results

### 6.1. Experimental protocol

Our experiments operate on 19 Mel-Frequency Cepstral Coefficients (plus energy) augmented with 19 first ($\Delta$) and 11 second ($\Delta\Delta$) derivatives. A mean and variance normalization (MVN) technique is applied on the MFCC features estimated using the

speech portion of the audio file. The low-energy frames (corresponding mainly to silence) are removed.
A gender-dependent 512 diagonal component GMM-UBM (male model) and a total variability matrix of low rank 400 are estimated using 15660 utterances corresponding to 1147 male speakers (using NIST SRE 2004, 2005, 2006 and Switchboard data). The LIA_SpkDet package of the LIA_RAL/ALIZE toolkit is used for the estimation of the total variability matrix and the i-vectors extraction. The used algorithms are described in [31]. Finally, a PLDA model is trained using 22264 sessions corresponding to 1857 speakers (1147 male + 710 female). The optimal eigenvoice rank, computed on the development set, is equal to 100 and the eigenchannel matrix is kept full-rank (400). PLDA is preceded by 2 iterations of LW-normalization [6].

### 6.2. The $C_{llr}$ performance measure

The $C_{llr}$ is a performance measure for SR systems and its formula is given by Equation 11.

$$C_{llr} = \frac{1}{2}\Big(\frac{\sum log_2(1+\frac{1}{LR})}{N_{tar}} + \frac{\sum log_2(1+LR)}{N_{non}}\Big) \quad (11)$$

Where $LR$ is system score, $N_{tar}$ and $N_{non}$ are respectively the number of target and non-target trials.

$C_{llr}$ -largely used in forensic voice comparison- is a loss in terms of likelihood ratio discrimination power. It does not require threshold and hard decisions like *equal error rate* (EER) [38]. $C_{llr}$ has the meaning of a cost or a loss: lower is the $C_{llr}$, better is the performance. We use the calibrated $C_{llr}$ and the minimum value of the $C_{llr}$ (denoted respectively $C_{llr}^{cal}$ and $C_{llr}^{min}$). $C_{llr}^{cal}$ involves calibration loss while $C_{llr}^{min}$ contains only discrimination loss. We can judge the quality of the calibration $Q_{cal}$ (i.e., the mapping from score to *log-likelihood-ratio* (LLR) which is actually present in the detector) by:

$$Q_{cal} = C_{llr}^{cal} - C_{llr}^{min}. \quad (12)$$

### 6.3. Application of I-MAP

We test the I-MAP denoising procedure described in Section 3 on the evaluation set. In this experiment, only long (more than 30s of speech) noisy ($SNR < 10dB$) test segments are used. For each test session, the algorithm described in Figure 1 is applied (6000 i-vectors are used to estimate $f(X)$ and 500 to estimate $f(N)$). Finally, the clean PLDA scoring is performed.

Table 1: Performance of I-MAP on the test set of SITW.

| Baseline | | | I-MAP | | |
|---|---|---|---|---|---|
| **EER** | **DCF** | $C_{llr}^{min}$ | **EER** | **DCF** | $C_{llr}^{min}$ |
| 12.69 | 0.9401 | 0.623 | 6.34 | 0.7481 | 0.324 |

It is clear that I-MAP improves significantly the performance of the SR system achieving 50% of relative EER improvement on noisy test data compared to the baseline system performance.

### 6.4. Homogeneity measure impact

Figure 2 presents for all the core-core evaluation trials $C_{llr}^{min}$ in function of $NHM$. In order to compute the $C_{llr}$ corresponding to a given $NHM$ value, we apply on the trials sorted by homogeneity values a 70000 (10% of all trials) sliding window, moved using a step of 70000 values. On each window, we compute the averaged $C_{llr}^{min}$ to be compared with the $NHM$ value (computed here as the median value on the window). A large correlation between the homogeneity values and the $C_{llr}^{min}$ is observed, with a quite consistent evolution from (HM=1208, $C_{llr}^{min}$=0.51) to (HM=9076, $C_{llr}^{min}$=0.27). This finding prove that *NHM* is able to predict the expected $C_{llr}^{min}$ based only on the two speech recordings of a given voice comparison trial. Consequently, this factor have to be taken into account for a reliable estimation of SR systems' abilities.

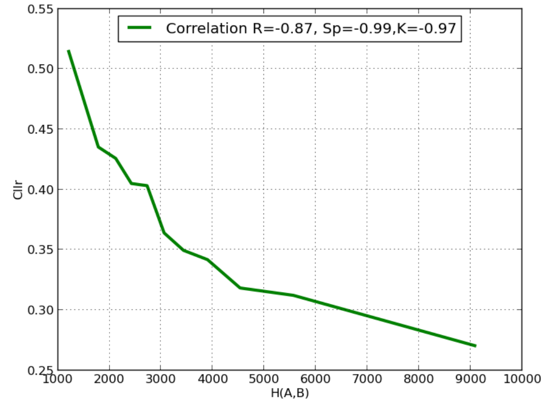

Figure 2: $NHM$ behavioral curve for the pooled condition (all the comparison tests taken together).

### 6.5. Using the discriminative classifier for scoring

In this experiment, the discriminative classifier described in Section 4 is used in the scoring phase in order to optimize the matricial parameters of the PLDA model based on the SITW development data. Table 2 compares its performance to a regular PLDA scoring.

Table 2: Performance of the discriminative classifier on the core-core evaluation data.

| Baseline | | | Discr. class. | | |
|---|---|---|---|---|---|
| **EER** | **DCF** | $C_{llr}^{min}$ | **EER** | **DCF** | $C_{llr}^{min}$ |
| 12.64 | 0.8442 | 0.428 | 11.93 | 0.8431 | 0.394 |

It is clear that this algorithm yields significant improvement when compared to a regular PLDA model in terms of EER and $C_{llr}^{min}$ which proves its efficiency in adverse test conditions and its interest as a generic optimization technique.

## 7. Conclusion

In this paper, we presented the speaker verification system developed in the LIA lab for the SITW (Speakers In The Wild) challenge. In order to improve the system performance, we tested new algorithms which operate on different levels. The I-MAP denoising procedure was used on trials involving noisy test data and showed 50% of relative EER improvement compared to a baseline system performance. Also, a new discriminative classifier was introduced as a tool to improve the matricial hyperparameters of the PLDA model. This system achieved significant improvement in recognition performance and was shown to be effective when used on adverse conditions (core-core task). Finally, the homogeneity measure was used to study the reliability of the SR system outputs based on an information-theoretic approach.

# 8. References

[1] P. Kenny, "Joint factor analysis of speaker and session variability: Theory and algorithms," *CRIM, Montreal,(Report) CRIM-06/08-13*, 2005.

[2] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 4, pp. 788–798, 2011.

[3] J. H. Hansen and T. Hasan, "Speaker recognition by machines and humans: A tutorial review," *Signal Processing Magazine, IEEE*, vol. 32, no. 6, pp. 74–99, 2015.

[4] T. Kinnunen and H. Li, "An overview of text-independent speaker recognition: From features to supervectors," *Speech communication*, vol. 52, no. 1, pp. 12–40, 2010.

[5] D. Garcia-Romero and C. Y. Espy-Wilson, "Analysis of i-vector length normalization in speaker recognition systems." in *Interspeech*, 2011, pp. 249–252.

[6] P.-M. Bousquet, A. Larcher, D. Matrouf, J.-F. Bonastre, and O. Plchot, "Variance-spectra based normalization for i-vector standard and probabilistic linear discriminant analysis." in *Odyssey*, 2012, pp. 157–164.

[7] A. Kanagasundaram, D. Dean, S. Sridharan, M. McLaren, and R. Vogt, "I-vector based speaker recognition using advanced channel compensation techniques," *Computer Speech & Language*, vol. 28, no. 1, pp. 121–140, 2014.

[8] P.-M. Bousquet, D. Matrouf, and J.-F. Bonastre, "Intersession compensation and scoring methods in the i-vectors space for speaker recognition." in *INTERSPEECH*, 2011, pp. 485–488.

[9] S. J. Prince and J. H. Elder, "Probabilistic linear discriminant analysis for inferences about identity," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE, 2007, pp. 1–8.

[10] P. Kenny, "Bayesian speaker verification with heavy-tailed priors." in *Odyssey*, 2010, p. 14.

[11] A. El-Solh, A. Cuhadar, and R. Goubran, "Evaluation of speech enhancement techniques for speaker identification in noisy environments," in *Multimedia Workshops, 2007. ISMW'07. Ninth IEEE International Symposium on*. IEEE, 2007, pp. 235–239.

[12] S. O. Sadjadi and J. H. Hansen, "Assessment of single-channel speech enhancement techniques for speaker identification under mismatched conditions." in *INTERSPEECH*, 2010, pp. 2138–2141.

[13] S. Liu, Y. Zou, and H. Ning, "Nonnegative matrix factorization based noise robust speaker verification," in *Signal and Information Processing (ChinaSIP), 2015 IEEE China Summit and International Conference on*. IEEE, 2015, pp. 35–39.

[14] S. Sarkar and K. Sreenivasa Rao, "Stochastic feature compensation methods for speaker verification in noisy environments," *Applied Soft Computing*, vol. 19, pp. 198–214, 2014.

[15] Y. Lei, L. Burget, and N. Scheffer, "A noise robust i-vector extractor using vector taylor series for speaker recognition," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 6788–6791.

[16] C. Yu, G. Liu, S. Hahm, and J. H. Hansen, "Uncertainty propagation in front end factor analysis for noise robust speaker recognition," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 4017–4021.

[17] D. Ribas, E. Vincent, and J. R. Calvo, "Uncertainty propagation for noise robust speaker recognition: the case of nist-sre," in *Interspeech 2015*, 2015.

[18] Y. Lei, L. Burget, L. Ferrer, M. Graciarena, and N. Scheffer, "Towards noise-robust speaker recognition using probabilistic linear discriminant analysis," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 4253–4256.

[19] N. Li and M.-W. Mak, "Snr-invariant plda modeling for robust speaker verification," *Proc. Interspeech'15*, 2015.

[20] S. Du, X. Xiao, and E. S. Chng, "Dnn feature compensation for noise robust speaker verification," in *Signal and Information Processing (ChinaSIP), 2015 IEEE China Summit and International Conference on*. IEEE, 2015, pp. 871–875.

[21] M. McLaren, Y. Lei, N. Scheffer, and L. Ferrer, "Application of convolutional neural networks to speaker recognition in noisy conditions." in *INTERSPEECH*, 2014, pp. 686–690.

[22] Y. Lei, L. Ferrer, M. McLaren *et al.*, "A novel scheme for speaker recognition using a phonetically-aware deep neural network," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 1695–1699.

[23] D. C. A. L. Mitchell McLaren, Luciana Ferrer, "The speakers in the wild (sitw) speaker recognition database," in *Submitted to Interspeech 2016.*

[24] W. Ben Kheder, D. Matrouf, J.-F. Bonastre, M. Ajili, and P.-M. Bousquet, "Additive noise compensation in the i-vector space for speaker recognition," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 4190–4194.

[25] W. B. Kheder, D. Matrouf, P.-M. Bousquet, J.-F. Bonastre, and M. Ajili, "Robust speaker recognition using map estimation of additive noise in i-vectors space," in *Statistical Language and Speech Processing*. Springer, 2014, pp. 97–107.

[26] D. Matrouf, W. Ben Kheder, P. Bousquet, M. Ajili, and J. Bonastre, "Dealing with additive noise in speaker recognition systems based on i-vector approach," in *Signal Processing Conference (EUSIPCO), 2015 23rd European*. IEEE, 2015, pp. 2092–2096.

[27] M. Ajili, J.-F. Bonastre, S. Rossato, J. Kahn, and I. Lapidot, "An information theory based data-homogeneity measure for voice comparison," in *Interspeech 2015*, 2015.

[28] M. "Ajili, J.-F. Bonastre, S. Rossato, J. Kahn, and I. Lapidot, "Homogeneity measure for forensic voice comparison: A step forward reliability," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. Springer, 2015, pp. 135–142.

[29] M. Ajili, J.-F. Bonastre, S. Rossato, and J. Kahn, "Fabiole, a speech database for forensic speaker comparison," *International Conference on Language Resources, Evaluation and Corpora*, 2016.

[30] "The NIST year 2008 speaker recognition evaluation plan," http://www.itl.nist.gov/iad/mig//tests/sre/2008/, 2008, [Online; accessed 15-May-2014].

[31] D. Matrouf, N. Scheffer, B. G. Fauve, and J.-F. Bonastre, "A straightforward and efficient implementation of the factor analysis model for speaker verification." in *INTERSPEECH*, 2007, pp. 1242–1245.

[32] N. Brümmer and E. De Villiers, "The speaker partitioning problem." in *Odyssey*, 2010, p. 34.

[33] P. Kenny, "Bayesian speaker verification with heavy-tailed priors." in *Odyssey*, 2010, p. 14.

[34] P.-M. Bousquet and J.-F. Bonastre, "Constrained discriminative speaker verification specific to normalized i-vectors," in *Odyssey*, 2016.

[35] T. Okada and S. Tomita, "An optimal orthonormal system for discriminant analysis," *Pattern Recognition*, vol. 18, no. 2, pp. 139–144, 1985.

[36] J. Wang, Y. Xu, D. Zhang, and J. You, "An efficient method for computing orthogonal discriminant vectors," vol. 73, no. 10-12, pp. 2168 – 2176, 2010.

[37] W. Y. Zhao, "Discriminant component analysis for face recognition," *Pattern Recognition*, vol. 2, pp. 818–821, 2000.

[38] N. Brümmer and J. Du Preez, "Application-independent evaluation of speaker detection," *Computer Speech & Language*, vol. 20, no. 2, pp. 230–275, 2006.