The 5th International Conference on Electrical Engineering – Boumerdes (ICEE-B)
October 29-31, 2017, Boumerdes, Algeria.

Multi-condition Gaussian Probabilistic Linear Discriminate Analysis in Automatic Speaker Recognition

Mourad DJELLAB[1] and Abderrahmane AMROUCHE[2]
SCSPL, Faculty of Electronics and Computer Science, USTHB
P.O Box 32, Bab Ezzouar, Algiers, Algeria
[1]mourad.djellab@northumbria.ac.uk

Noureddine MEHALLEGUE
Research and Development Centre (CRD-GN), Bouchaoui, Algeria
nmehallegue01@qub.ac.uk

Ahmed BOURIDANE
Department of Computer Science and Digital Technologies,
Northumbria University, Newcastle Upon Tyne, NE1 8ST, UK
ahmed.bouridane@northumbria.ac.uk

Abstract— **Automatic Speaker Recognition Systems (ASRSs) are being used in forensic investigations to help experts to appreciate the measure of confidence related to real vocal comparison cases. Accuracy is an important aspect in such kind of applications where the existing recognition approaches are known to be sensitive to recording conditions mismatch, the utterances lengths mismatch, and the linguistic variation effects. This paper proposes a recognition strategy based on the use of Multi-condition Gaussian Probabilistic Linear Discriminate Analysis (MGPLDA) to compensate the effects of this type of perturbation sources. Experiments using an Algerian Speaker Recognition Database collected in different recording conditions have shown that the proposed approach successfully improves the system performances in terms of discrimination and calibration quality.**

Keywords-component; Speaker recognition; multi-conditions scenarios; GPLDA; MGPLDA;

I. INTRODUCTION

Voice recognition has received a lot of attention in biometrics during the last decade, particularly after the introduction of automatic speaker recognition systems (ASRS) in forensic laboratories where meaningfully interpreted identification scores are required in addition to other necessary appreciations related to the output of the system confidence measure [1][2][3]. However, it has been noted that guaranteeing the accuracy and the robustness of the recognition approaches actually used in this domain in real life applications is a challenging task. One of the most important topics in this context is related to the effect of using data different to that used during the development processes of these systems in operating mode [1]. The mismatch between the development data used to train the recognition model and the evaluation data encountered during operating modes, results generally in some performance losses, particularly in terms of accuracy and calibration quality [4].

To compensate the deficiency in calibration quality for example, researchers in [2] have suggested the use of separate calibration models trained for each particular recording condition in order to optimize the calibration for each of them; however, this requires prior knowledge of the trial conditions. Thus, conditions not considered during training of the calibration model cannot be handled in this case because it is not realistic to anticipate all the potential situations that the system is going to face during deployment. Furthermore, the benefit of using separate calibration models in ASRS as proposed in [2] is limited to the improvement of the calibration quality.

In this work, a recognition strategy is proposed based on the use of Multi-condition Gaussian Probabilistic Linear Discriminate Analysis (MGPLDA) in ASRS. The mismatch between the ASRS development data and that encountered during its operating modes does not affect only the calibration quality, but it also affects the accuracy of the ASRS. Therefore, it would be wiser if the proposed approach could jointly improve the accuracy of the ASRS while still compensating the calibration loss related to the data recording conditions mismatch. More importantly, the proposed approach should be able to approximate the test recording conditions with the closest development condition and produce a score no worse than what the baseline system would produce. This makes the global accuracy of the system better than the baseline in the expected conditions and not worse (or better) in unanticipated scenarios which is more realistic in real applications.

The main purpose of this work is to investigate the advantages of using the MGPLDA to improve the calibration quality and the discrimination of ASRS in mismatch recording conditions. It explores how the linguistic variation and the recording conditions mismatch affect the target and non-target score distributions, the impact of this effect on the accuracy and the calibration quality and the possibility to compensate the performance losses based on using the MGPLDA approach.

II. RECORDING CONDITIONS AND LINGUISTIC VARIATION MISMATCH EFFECTS

The evaluation of the accuracy of an ASRS is not an easy task due to the difficulty to answer some questions related to this topic such as the size of the evaluation data and the number of target and non-target trials to be considered in the experiment [5]. Moreover, it is relatively easy to model the inter-speaker variability using a large variety of settings (channels, languages etc.) but it is challenging to estimate the intra-speaker variability since the recording voices being compared for each speaker may differ greatly from one case to

978-1-5386-0686-5/17/$31.00©2017

another [4]. A common solution to obtain good inter and intra-speaker variability estimations is to use a pool of reference speakers each having a large number of recordings collected in different case-specific settings. However, this is not always the case in evaluation corpuses [4].

In reality, one of the main challenge in ASRS is related to the target and non-target scores distributions and how they reflect some characteristics of the development and evaluation data such as matched (tel-tel), mismatched (tel-mic) recording conditions [4], language (languages, dialects and accents), and degradation (clean, noisy, reverberated). Currently, it is well known that using different sets of development data with different recording conditions extremely mismatched, lead to different target and non-target scores distributions with the fact that the distance between the respective target scores distribution and the non- target scores distribution depends on each set of development data recording condition. For example, if we consider different languages as the extreme mismatch condition for dialect or regional accents variation, it is clear that the effect of linguistic variation on the non-target scores distribution is significant. Fig.1 shows that the distance between the target and non-target scores distributions would be minimal (d1) for extreme matched conditions (same accent) and increase progressively according to the importance of linguistic variations to reach its maximal value (d2) for the extreme mismatched conditions (different languages).

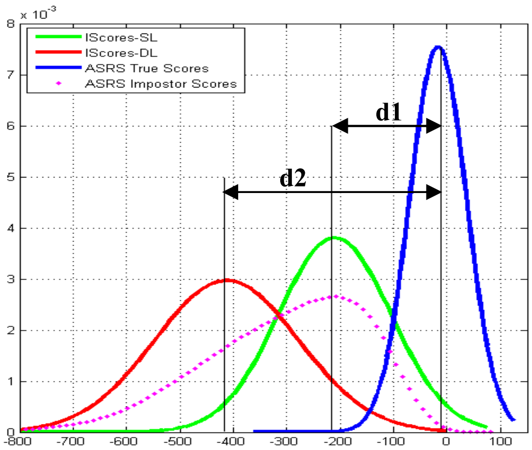

Figure 1: Effect of language mismatch (linguistic variations) on the non-target scores distribution

From this point of view, it can be concluded that guaranteeing the accuracy and the robustness of generative recognition approaches in real life applications is a challenging task. Thus, obtaining a non-linear calibration function, for example, for which the optimal decision threshold computed for each pair of enrolment and testing is independent of the trial conditions depends on the availability of datasets that closely represent the whole acoustic conditions expected to be encountered during end use of the system, this task is challenging and often not possible to achieve particularly in forensic applications. This conclusion is also valid for a Logistic Regression calibration technique when a single model is trained using all development data [8][9]. In this case, the approach optimizes the calibration process globally for all conditions in the development data which might not be optimal when the performance is computed for each condition separately, especially for conditions not considered during the training phase [2].

III. RECOGNITION USING THE MGPLDA

Many researchers consider that the most successful scheme for use in speaker recognition should combine the Total Variability space (TV) with LDA, and the PLDA [7][8][9][10]. In fact, LDA is applied for dimensionality and within classes' variability reduction and the PLDA is used at the verification stage to perform classification.

Motivated by the results reported in [9][11] regarding the efficiency of using multi-condition GPLDA models especially with mismatch data conditions in face recognition and the suggestions in [2] related to the possibility of using separate calibration model to overcome calibration loss, we propose in this paper two variants of MGPLDA Models: i) Hard-Decision MGPLDA and ii) the fusion of individual scores obtained from different GPLDA components using the logistic regression.

A. Hard-Decision MGPLDA (HD-MGPLDA)Models

One common assumption in conventional GPLDA model is that it ignores the i-vector extraction process and considers the i-vector as an observed variable following the generative model described in equation (1), where the model parameters are obtained from a large collection of development data using an EM algorithm as in [12].

$$\mathrm{n}_{ij} = \mu + Vy_i + \varepsilon_{ij} \quad (1)$$

where: $\mathrm{n_{ij}} \in \mathbb{R}^{d\,x\,1}$ is the extracted i-vector (j) from the utterance of speaker (i), μ is the data global mean, $\mathrm{y_i} \sim \aleph(0, \mathrm{I})$, and $\varepsilon_{\mathrm{ij}} \sim \aleph(0, \Sigma^{-1})$. V is a matrix describing the between- and within individual variability subspaces and Σ is the covariance matrix.

This is not the case of multi-condition setup where there is an access to K versions of the development data, and therefore it is possible to estimate a collection of GPLDA models (each model reflects the particular settings of its subset of training data) [13]–[15].

In HD-MGPLDA Models, we assume that the i-vectors $\{1 \ldots j\}$ of speaker (i) are generated using the same latent identity variable (y) and the same hyperparameters only if these i-vectors belong to the same subset of conditions (the same Gaussian component $\mathrm{G_k}$in our work). Otherwise, if the i-vectors of the speaker (i) belong to different Gaussian components, they are considered to be generated using different latent identity variable ($\mathrm{y_k}$) and with different hyper-parameters dependent of the subset of conditions (in this variant we admit k latent factors for each speaker, one for each mixture). This results in a generative model in the form:

$$\mathrm{n_{ij}} = \mu_\mathrm{k} + \mathrm{V_k y_{ik}} + \varepsilon_{\mathrm{ijk}}, \text{If } \mathrm{n_{ij}} \text{ belong to } \mathrm{G_k} \quad (2)$$

During verification phase (Fig.2), the GMM with the highest likelihood ratio determines which GPLDA component should be used for the scoring of the test i-vector (n). Then, to obtain the score between two i-vectors $\mathrm{n_1}$ and $\mathrm{n_2}$, two alternative hypotheses are considered: i) $\mathrm{H_s}$: both i-vectors share the same speaker identity latent variable (y), and ii) H_d: the i-vectors n_1 and n_2 are generated using different identity variables. The verification score can be computed in the same way as baseline GPLDA model [16]:

$$\mathrm{score} = \log \frac{P(n_1, n_2 | H_s)}{P(n_1 | H_d) P(n_2 | H_d)} \quad (3)$$

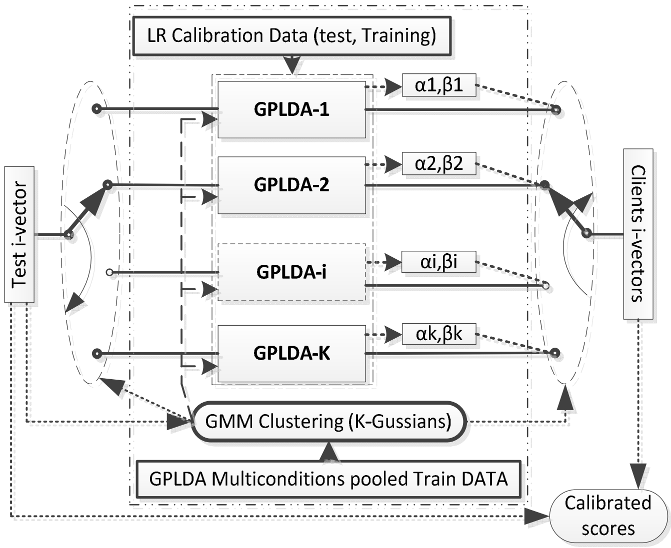

Figure 2: The dataflow of the HD-GPLDA method

B. *Fusion of Multi-condition GPLDA (F-MGPLDA) Models*

Two distinguishable alternatives are generally considered in this variant. One is to assume that each condition is independent of each other in order to obtain hyper-parameters as in the basic GPLDA model (independent GPLDA components). In this case, the i-vectors of the same speaker are assumed to be generated with the same identity variable (y) but using different hyper-parameters, only if all the i-vectors belong to the same subset of conditions. The i-vectors belonging to different subsets are considered to be generated using different latent identity variables (y_k) and different hyperparameters.

The second Multi-condition GPLDA alternative is the Tied-PLDA model where each speaker model is represented by k i-vectors (K models: $\{n_i^M\}_{i=1}^K$). In this case, the i-vectors of speaker (i) are supposed to be generated using the same latent identity variable (y), but the hyper-parameters of the subsets are different [12]. This results in the following generative model:

$$n' = \mu' + V'y + \varepsilon' \quad (4)$$

where:

$n' = [n_1^t, \dots n_k^t]^t$, $\mu' = [\mu_1^t, \dots \mu_k^t]^t$, $V' = [V_1^t, \dots V_k^t]^t$ and $\varepsilon' = [\varepsilon_1^t, \dots \varepsilon_k^t]^t$

In both cases: dependent and totally independent GPLDA components, the final score is obtained as a convex mixture of a collection of K scores s_i according to the mixing weights w_i, as shown in Fig.3.

$$score = \sum_{i=1}^{k} w_i\, s_i \quad (5)$$

The mixing weights w_i can be estimated using different techniques chosen according to the GPLDA model assumptions (totally independent or dependent). In our experimental section, we consider only the first scenario where the multi-class linear logistic regression fusion technique [17] may be employed to estimate w_i. In this case, the individual scores s_i are combined according to the same procedure used for the fusion of multiple recognizers with linear logistic regression [18]. For every trail t, the fused score is obtained according to:

$$\overrightarrow{\mathrm{score}}(t) = \sum_{i=1}^{K} \alpha_i\, \vec{s}_i(t) + \vec{\beta} \quad (6)$$

The parameters ($\alpha_1, \alpha_2 \dots, \alpha_K, \vec{\beta}$) are obtained during the development phase using a calibration /fusion training data.

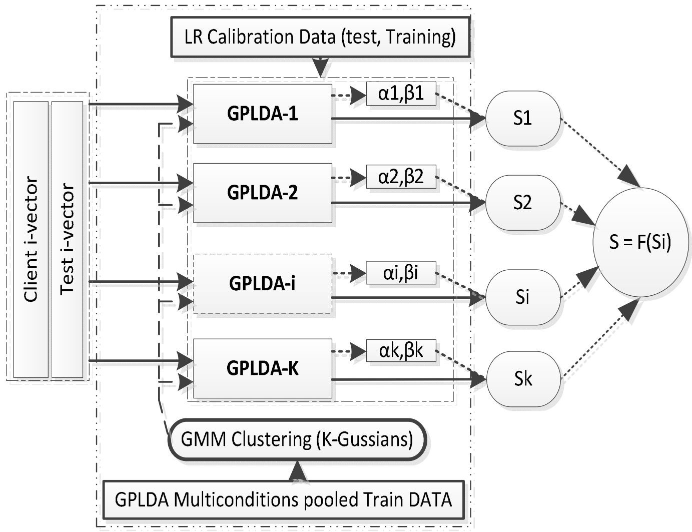

Figure 3: The dataflow of the F-GPLDA method

IV. SCORES SHIFTING BETWEEN GPLDA COMPONENTS

In the mathematical formulation of speaker recognition PLDA based systems, recognition scores are computed as likelihood-ratios. This is the case of MGPLDA models, however, it has been found that scoring the same evaluation data with separate GPLDA components produces scores at different ranges (scores shifting between GPLDA components) [19].

One way to combine the scores in this situation is to associate a dependent calibration model to each GPLDA component as described in Fig.2 and Fig.3. The parameters of the linear calibration model $f(\alpha_k, \beta_k)$ of the k^{th} GPLDA component, are obtained using a dataset different to that used to train the multi-condition GPLDA models and according to the same procedure used to train conventional linear calibration models with logistic regression [18].

In the proposed system, the score obtained for each trail is calculated using the GPLDA component trained using the closest development data condition and calibrated using the k^{th} GPLDA correspondent calibration model $f(\alpha_k, \beta_k)$.

V. EXPERIMENTAL SETUP

To conduct this experiment, we used two corpora of speech resources and three different configurations of an i-vectors speaker recognizer (baseline system and both variants of MGPLDA previously mentioned in section III).

The Algerian Modern Colloquial Arabic Speech Corpus (AMCASC): the particularities of the AMCASC is related to the large variety of settings characterizing the quality of their data (channels, environment noise, utterance length, amplitude clipping etc.) which makes it similar to that encountered in forensic applications. Moreover, the technical quality of the data is not the only particularity which should be considered in this corpus. The use of code switching is another factor which makes the speech utterances being compared differ greatly from one case to another [20].

MOBIO: contains a bi-model data (face/speaker) collected using two types of mobile devices: mobile phones and laptop computers, from 152 peoples (100 males and 52 females) native and non-native English speakers. More technical details about the MOBIO database can be found in [21].

A. Recognition approach

Towards the end of the last decade, the i-vector has become a reliable indicator for the current state-of-the-art in speaker verification applications [22]. This compact representation provides a way for reducing large-dimensional input data to a low-dimensional space while retaining most of the speaker relevant information [23]. The key idea of this approach is to combine both speaker and channel/session variability in the same space. Thus, speaker and session-dependent super-vectors of concatenated GMM can be written as [22]:

$$M = m + Tw \tag{7}$$

where m is the Universal Background Model (UBM) mean super-vector, T is a low rank matrix that defines the low-dimensional space, and w is a standard-normally distributed latent variable representing the identity vector.

In practice, a large database is used to train the UBM (1024) and the total variability subspace matrix T (size 400) which are necessary to calculate the i-vectors. Then, an LDA technique is applied to reduce the i-vectors dimensionality to 200, 250, 300 and 350. In order, to make the development and trial i-vector distributions closer and more Gaussian shaped, the i-vector length normalization technique, proposed in [16], have been used to avoid dataset shift by normalizing each i-vector by its magnitude. Finally, a GPLDA model and both proposed variants of MGPLDA models previously described have been employed to calculate the recognition scores.

B. Results and discussion

According to the purpose of this work, the results of the experimental evaluation are given in two subsections:

1) Recognition discrimination:

To evaluate the effectiveness of the proposed methods in terms of accuracy, we made use of the DET curve, the Equal Error Rate (EER) and the normalized minimum Detection Cost Function (min-DCF) as the metrics. The performance of baseline GPLDA, HD-MGPLDA and F-MGPLDA systems are shown in Table 1 and Fig.4.

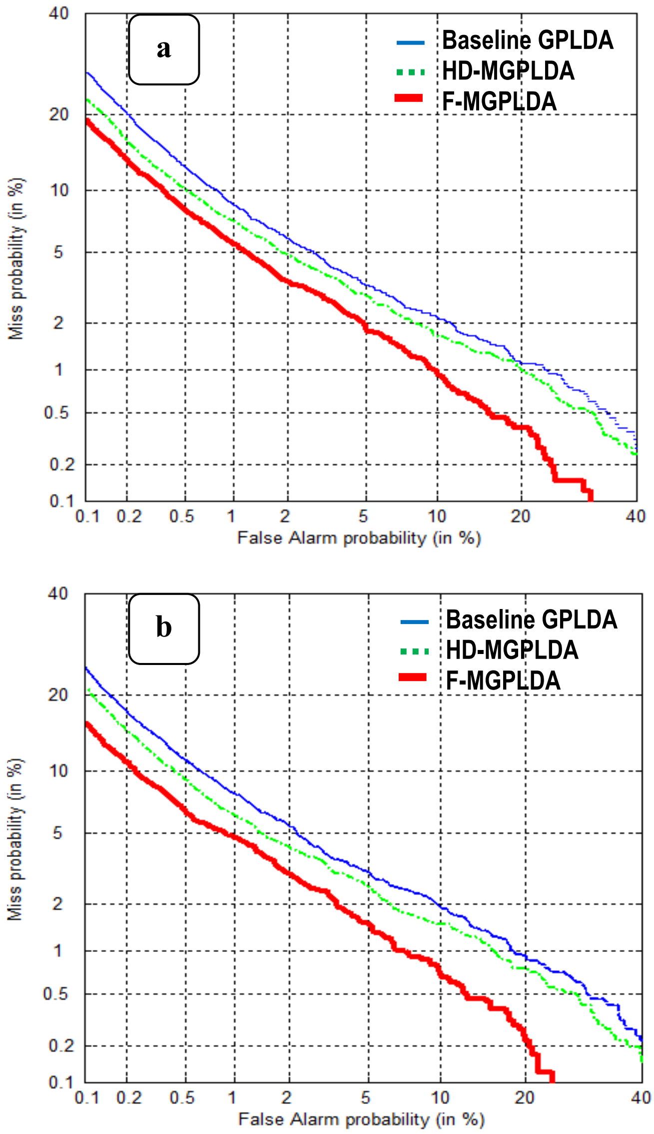

Figure 4: Recognition performances using DET curves
a) LDA_{DIM} =200, N_{GPLDA} = 2, **b)** LDA_{DIM} =350, N_{GPLDA} = 2

Table 1 summarizes the speaker recognition performance of the proposed methods and the baseline system in different configurations: Methods (HD-MGPLDA (M1), F-MGPLDA (M2)), number of GPLDA components (N_{GPLDA}) and the LDA dimension (dim_{LDA}). As expected, it can be clearly seen in Table 1 that some configurations of the proposed methods improve the recognition accuracy of the ASRS where the GPLDA-ivectors baseline system is used as a reference point. The obtained relative improvements compared to the best results obtained with the GPLDA-ivectors baseline system ($dim_{LDA} = 350$), in the configurations: (M1, 2, 300), (M1, 2, 350), (M1, 3, 350), (M2, 2, 250), (M2, 2, 300), (M2, 2, 350), are 12.36%, 17.41%, 3.93%, 18.26%, 27.25% and 28.08%, respectively.

TABLE 1: GPLDA AND MGPLDA SYSTEMS PERFORMANCES ACCORDING TO THE LDA DIMENSION

LDA Dimension	Baseline GPLDA	HD-MGPLDA		F-GPLDA	
		$N_{GPLDA} = 2$	$N_{GPLDA} = 3$	$N_{GPLDA} = 2$	$N_{GPLDA} = 3$
200	4.06	3.72	3.91	3.22	4.24
	[0.0791]	[0.0704]	[0.0764]	[0.0621]	[0.0814]
250	3.9	3.28	3.77	2.91	4.05
	[0.0754]	[0.0639]	[0.0726]	[0.0544]	[0.0778]
300	3.69	3.12	3.51	2.59	3.94
	[0.0704]	[0.0599]	[0.0693]	[0.0506]	[0.0766]
350	3.56	2.94	3.42	2.56	3.95
	[0.0697]	[0.0580]	[0.0671]	[0.0505]	[0.0773]

The results obtained show also that, both HD-GPLDA and F-GPLDA systems produce the best results with two GPLDA components where different levels of EER reduction compared to the baseline PLDA system are obtained. This is a bit surprising result because in this case, we have only two physical clusters and it is generally admitted that the more clusters we use, the better results we should obtain. However, it is worth noting that our results are in perfect agreement with those reported in [7] on the use of unsupervised learning of PLDA mixture models (homogeneous i-vector extractor) for speaker verification in the sense that no gain is obtained with a number of mixture $N_{GPLDA} > 2$. Therefore, this result might be related to the quality of the estimated MGPLDA model, which is too sensitive to the quantity of data available to train each GPLDA component. It is important to note that the F-GPLDA system outperforms the other used methods (HD-GPLDA and baseline PLDA systems) and they achieve peak performance when $N_{GPLDA} = 2$ and $\dim_{LDA} \geq 350$.

Furthermore, the results obtained show that even with lower LDA dimensions ($\dim_{LDA}$), the MGPLDA models outperform the baseline GPLDA system in its best configuration ($\dim_{LDA} = 350$ in our experiment). As example, the EERs obtained using the F-GPLDA with $\dim_{LDA} = 200$ and 250, in addition to that obtained using H-GPLDA with a $\dim_{LDA} = 250$, are: 3.22%, 2.91% and 3.28%, respectively. These results are better than that obtained using the baseline PLDA system (3.56%) with a $\dim_{LDA} = 350$. The advantage of MGPLDA models from this point of view might help to compensate certain other disadvantages of this approach, particularly, those related to the necessity to train multiple GPLDA models, which increases computation complexity during the training and the recognition stages.

2) *Calibration quality*

In this subsection, we evaluate our proposed ASRS in terms of calibration quality. Therefore, we have used the normalized Bayes error-rate plot which is a convenient way of visualizing calibration performances over a wide range of operating points and gives the possibility to compare the calibration quality of the proposed method with baseline systems, in different operating points [24]. The importance of this metric is that if the ASRS is evaluated at just one specific application, as represented by just one point on the horizontal axis, then the measurement at that point would not necessarily give a good indication of usefulness at some other point (application) along the axis [24].

In addition, both the minimum Detection Cost function (MinDCF) and the actual Detection Cost Function (ActDCF) are measured. The calibration error is defined as the difference between ActDCF and MinDCF for each recognition approach [25].

Fig.5 shows the normalized Bayes error rate plots of the proposed methods and the baseline GPLDA system in different operating points. It is clearly notable that both proposed methods perform better in terms of calibration quality than the baseline GPLDA system almost everywhere. The only region where there is a small degradation and the baseline GPLDA system outperforms the proposed methods is approximately between [-7, -2] ($-7 \leq \text{logit}\,\pi \leq -2$).

Table 2 confirms the results previously reported and shows the best calibration quality obtained in our experiment using the F-GPLDA method where a lower LDA dimension ($\dim_{LDA} = 200$) is employed. Moreover, it can be clearly noted that the worst result obtained using the F-GPLDA method (actDCF- minDCF = 16.81% where $\dim_{LDA} = 350$) outperforms the best result obtained using the baseline GPLDA system (ActDCF - minDCF = 20.63% where $\dim_{LDA} = 200$).

The results show also that the improvements obtained in terms of recognition discrimination, when the subspace dimension ($\dim_{LDA}$) increase, are associated with a small degradation in term of calibration quality. However, it is clearly notable that the proposed methods outperform the baseline GPLDA system in term of recognition discrimination and calibration quality even in worst configurations;

Based on the analyses of the previous results, the idea of using MGPLDA models in ASRS is turned out to be very useful in the case of mismatched recording conditions.

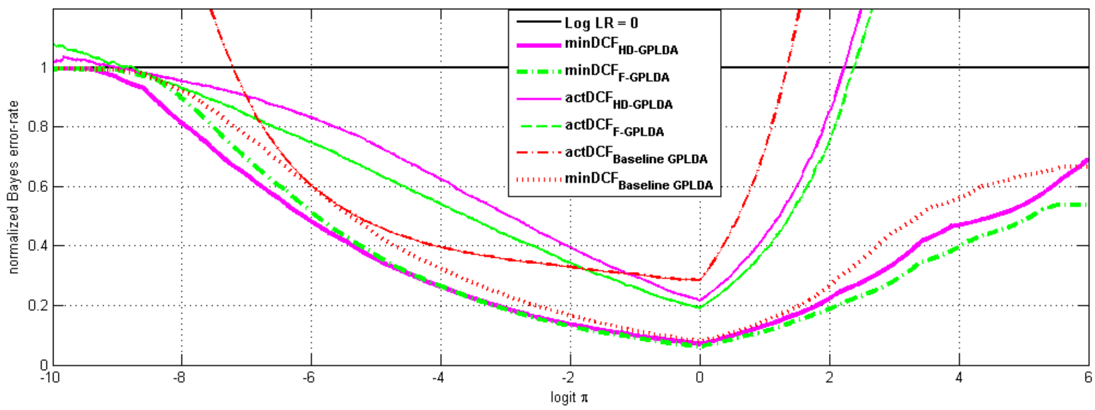

Figure 5: Normalized Bayes error rate plots of the evaluated methods ($N_{GPLDA} = 2$, LDA dimension = 200)

TABLE 2: EVALUATION OF THE CALIBRATION QUALITY OF THE USED METHODS ($\mathbf{N_{GPLDA}} = \mathbf{2}$)

		200	**250**	**300**	**350**
Baseline GPLDA	EER %	**4.06**	**3.90**	**3.69**	**3.56**
	minDCF %	7.91	7.54	7.04	6.97
	actDCF %	28.54	29.09	29.54	29.14
	(actDCF- minDCF) %	**20.63**	**21.55**	**22.50**	**22.17**
HD-GPLDA	EER %	**3.72**	**3.28**	**3.06**	**2.85**
	minDCF %	7.04	6.39	5.94	5.57
	actDCF %	21.57	23.75	24.65	27.75
	(actDCF- minDCF) %	**14.53**	**17.36**	**18.71**	**22.18**
F-GPLDA	EER %	**3.22**	**2.91**	**2.59**	**2.56**
	minDCF %	6.21	5.44	5.06	5.01
	actDCF %	19.12	20.24	20.11	21.82
	(actDCF- minDCF) %	**12.91**	**14.08**	**15.05**	**16.81**

VI. CONCLUSION

This work investigates the use of Multi-condition Gaussian probabilistic Linear Discrimant Analysis (MGPLDA) in mismatched recording condition scenarios, particularly the benefits of using this classification methodology in terms of discrimination and calibration quality.

Although the main limitation of multi-condition GPLDA models is related to the necessity to train multiple PLDA components, which increases computation complexity during the training and recognition stages, it has been shown through intensive experiments that this approach performs better with recording condition mismatch than baseline GPLDA systems. The obtained results using AMCASC, which is a real life speech corpus, shows that the proposed approach offer better discrimination and overall calibration performance than the baseline GPLDA system in ASRS. This conclusion is based on a series of evaluations using a real life speech data similar to that encountered in forensics. However, regarding the importance of the accuracy in such kind of applications, future research should cover more simulations with other recording conditions such as reverberation in order to strength the proposed methodology.

REFERENCES

[1] T. Ali, L. Spreeuwers, R. Veldhuis, and D. Meuwly, "Biometric evidence evaluation: an empirical assessment of the effect of different training data," IET Biometrics, vol. 3, no. 4, pp. 335–346, 2014.

[2] M. McLaren, A. Lawson, L. Ferrer, N. Scheffer, and Y. Lei, "Trial-Based Calibration for Speaker Recognition in Unseen Conditions," in The Speaker and Language Recognition Workshop (Odyssey), 2014, no. June, pp. 19–25.

[3] M. I. Mandasari, M. Günther, R. Wallace, R. Saeidi, S. Marcel, and D. A. van Leeuwen, "Score calibration in face recognition," Biometrics, IET, vol. 3, no. 4, pp. 246–256, 2014.

[4] Y. A. Solewicz, G. Jardine, T. Becker, and S. Gfroerer, "Estimated Intra-Speaker Variability Boundaries in Forensic Speaker Recognition Casework," in Proceedings of Biometric Technologies in Forensic Science (BTFS), 2013.

[5] D. A. Van Leeuwen and N. Brümmer, "An introduction to application-independent evaluation of speaker recognition systems," in Speaker classification I, Springer, 2007, pp. 330–353.

[6] M. I. Mandasari, R. Saeidi, M. McLaren, and D. a. Van Leeuwen, "Quality measure functions for calibration of speaker recognition systems in various duration conditions," IEEE Trans. Audio, Speech Lang. Process., vol. 21, no. 11, pp. 2425–2438, 2013.

[7] T. Pekhovsky and A. Sizov, "Comparison between supervised and unsupervised learning of probabilistic linear discriminant analysis mixture models for speaker verification," Pattern Recognit. Lett., vol. 34, no. 11, pp. 1307–1313, 2013.

[8] S. Cumani, O. Plchot, and P. Laface, "On the use of i–vector posterior distributions in Probabilistic Linear Discriminant Analysis," IEEE/ACM Trans. Audio, Speech, Lang. Process., vol. 22, no. 4, pp. 846–857, 2014.

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

[10] P. Rajan, A. Afanasyev, V. Hautamäki, and T. Kinnunen, "From single to multiple enrollment i-vectors: Practical PLDA scoring variants for speaker verification," Digit. Signal Process., vol. 31, pp. 93–101, 2014.

[11] S. J. D. Prince, P. Li, Y. Fu, U. Mohammed, and J. Elder, "Probabilistic models for inference about identity," IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 1, pp. 144–157, 2012.

[12] D. Garcia-Romero, X. Zhou, and C. Y. Espy-Wilson, "Multicondition training of Gaussian PLDA models in i-vector space for noise and reverberation robust speaker recognition," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012, pp. 4257–4260.

[13] N. Li and M.-W. Mak, "SNR-invariant PLDA modeling in nonparametric subspace for robust speaker verification," IEEE/ACM Trans. Audio, Speech, Lang. Process., vol. 23, no. 10, pp. 1648–1659, 2015.

[14] M.-W. Mak, X. Pang, and J.-T. Chien, "Mixture of PLDA for Noise Robust I-Vector Speaker Verification," IEEE/ACM Trans. Audio, Speech, Lang. Process., vol. 24, no. 1, pp. 130–142, 2016.

[15] M.-W. Mak, "SNR-Dependent Mixture of PLDA for Noise Robust Speaker Verification," in Fifteenth Annual Conference of the International Speech Communication Association, 2014.

[16] D. Garcia-Romero and C. Y. Espy-Wilson, "Analysis of i-vector Length Normalization in Speaker Recognition Systems.," in The Annual Conference of the International Speech Communication Association (Interspeech), 2011, pp. 249–252.

[17] N. Brümmer, "FoCal multi-class: Toolkit for evaluation, fusion and calibration of multi-class recognition Scores—Tutorial and user manual—," 2007.

[18] S. Pigeon, P. Druyts, and P. Verlinde, "Applying logistic regression to the fusion of the NIST'99 1-speaker submissions," Digit. Signal Process., vol. 10, no. 1, pp. 237–248, 2000.

[19] X. Pang and M.-W. Mak, "Fusion of SNR-dependent PLDA models for noise robust speaker verification," in 9th International Symposium on Chinese Spoken Language Processing (ISCSLP), 2014, pp. 619–623.

[20] M. Djellab, A. Amrouche, A. Bouridane, and N. Mehallegue, "Algerian Modern Colloquial Arabic Speech Corpus (AMCASC): regional accents recognition within complex socio-linguistic environmentsle," Lang. Resour. Eval., pp. 1–29, 2016.

[21] MOBIO, "MOBIO - Mobile Biometry," 2008. [Online]. Available: http://www.mobioproject.org/. [Accessed: 10-Mar-2016].

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

[23] L. Burget, O. Plchot, S. Cumani, O. Glembek, P. Matějka, and N. Brümmer, "Discriminatively trained probabilistic linear discriminant analysis for speaker verification," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2011, pp. 4832–4835.

[24] N. Brümmer, "Measuring, refining and calibrating speaker and language information extracted from speech," Stellenbosch, University of Stellenbosch, 2010.

[25] V. Hautamaki, T. Kinnunen, F. Sedlák, K. A. Lee, B. Ma, and H. Li, "Sparse classifier fusion for speaker verification," IEEE Trans. Audio. Speech. Lang. Processing, vol. 21, no. 8, pp. 1622–1631, 2013.