

## Component Score Weighting for GMM based Text-Independent Speaker Verification

Liang Lu<sup>2</sup>, Yuan Dong<sup>1,2</sup>, Xianyu Zhao<sup>1</sup>, Hao Yang<sup>2</sup>, Jian Zhao<sup>2</sup>, Haila Wang<sup>1</sup>

<sup>1</sup> France Telecom R&D Center (Beijing), Beijing, 100080, P. R. China

{yuan.dong, xianyu.zhao, haila.wang}@orange-ftgroup.com

<sup>2</sup> Beijing University of Posts and Telecommunications, Beijing, 100876, P. R. China

yuandong@bupt.edu.cn {luliang07, haoyang.hy, michaeljianzhao}@gmail.com

### Abstract

GMM/UBM framework is widely used in Automatic Speaker Verification (ASV), however, due to the insufficiency of the training data, both the hypothesized speaker and impostors are not well modeled, especially to some of the Gaussian component mixtures. Thus, the Gaussian mixtures in each GMM model have different discriminative capabilities, and the mismatch between testing and training data will also aggravate this situation. In this paper, we propose a novel approach, namely, Component Score Weighing (CSW), to reweight the Gaussian mixtures and highlight those which have high discriminative capability by post-processing the log-likelihood ratio (LLR). The original log-likelihood in GMM systems is assigned to each Gaussian component mixture, deriving two component score serials, which we called the dominant score serial and the residual score serial. A nonlinear score weighting function is then applied to reweigh those scores, respectively. Experiments on NIST 2006 SRE corpus show that, this approach achieves notable performance gains over our previous baseline system (about 12% relative improvement in minimum detection cost function (DCF) value).

### 1. Introduction

Over the past decade, Gaussian Mixture Models (GMMs) have become the dominant approach in text-independent speaker verification and identification [1]. In this approach, a speaker model is adapted by Maximum a Posteriori (MAP) form Universal Background Model (UBM), and verification in these GMM-UBM systems is carried out at frame level with an overall score for the whole utterance by averaging the likelihood ratios of each frame in the utterance.

This framework achieves good performance, but still has its problems. The training data for adapting speaker models is not sufficient, and usually comes from only a single audio session. Consequently, a speaker model will not only represent statistics of the speaker, but also the speech content, the handset type and the audio environment, etc. And the mismatch of training and testing condition is also one of the major reasons for detection errors. In addition, the log-likelihood ratio detection approach also has its limitations, just as pointed in [2], the likelihood score is achieved by frame-by-frame discrimination, whereas speaker verification is concerned with the discrimination of the whole utterance.

To resolve the problems mentioned above and improve speaker verification system performance, researchers mainly work in two areas, one is to perform the speaker verification in

utterance level, and several approaches are proposed in recent years, which are based on Support Vector Machines (SVMs) classifier [2, 3, 4]. Another one is to compensate the mismatch between training and testing data. These compensation techniques include widely used approaches such as RASTA filtering [5], feature mapping [6], and various score normalization techniques (Hnorm [7], Tnorm[8], ATnorm[9]), etc.

In this paper, we propose a novel approach, namely, Component Score Weighting (CSW), to alleviate two of the major reasons for decision errors in speaker verification, which are the insufficiency of the training data and the mismatch of the training and testing condition. We suppose that both of the situations will make the Gaussian mixtures in each GMM model have different discriminative capabilities, and the likelihood score of some Gaussian mixtures may have little contribution to the final recognition performance. Under this assumption, we first assign the log-likelihood score to each Gaussian component mixture, deriving two component score serials, which we called the dominant score serial and the residual score serial. A nonlinear score weighting function is then applied to the component score serials to reweight those scores. Experiments show that, by performing the nonlinear score weighting function, the final score will have better discriminative capability, and obvious system performance gains are obtained over our previous baseline system.

The paper is organized as follows: In section 2, we give a brief overview of classical GMM-UBM based speaker verification and log-likelihood ratio detection. In section 3, we extend the likelihood score in GMMs and describe our Component Score Weighting (CSW) approach in detail. Experimental results are given in section 4 and section 5 gives the conclusion and future work.

### 2. GMM System Description

In classical GMM-UBM based speaker verification systems, a general Universal Background Model (UBM) is trained using the EM algorithm on a larger quantity of exclusive speech, and the target speaker model is adapted from the gender specific UBM using MAP estimation. For a D-dimensional feature vector  $\vec{x}$ , the probability density of  $\vec{x}$  given a speaker model  $\lambda$ , which has  $M$  Gaussian mixtures, is defined as follows:

$$p(\vec{x}|\lambda) = \sum_{i=1}^M w_i p_i(\vec{x}). \quad (1)$$

The density is a weighted linear combination of  $M$  unimodal Gaussian densities  $p_i(\vec{x})$ , which is parameterized by a mean vector  $\vec{\mu}_i$  and a covariance matrix  $\Sigma_i$ . The speaker model can be characterized by  $\lambda = (w_i, \vec{\mu}_i, \Sigma_i), i = (1, \dots, M)$ .

Given a GMM model  $\lambda$ , the log-likelihood of the test utterance  $X = \{\vec{x}_1, \dots, \vec{x}_T\}$  is computed by:

$$\log P(X|\lambda) = \frac{1}{T} \sum_t \log p(\vec{x}_t|\lambda) \quad (2)$$

And the log-likelihood ratio (LLR) used for detection is as follows:

$$\Lambda(X) = \log p(X|\lambda_{spk}) - \log p(X|\lambda_{ubm}) \quad (3)$$

where  $\lambda_{ubm}$  is the UBM model, and  $\lambda_{spk}$  indicates the speaker model, respectively.  $\Lambda(X)$  is finally compared with the predefined threshold  $\theta$ , to decide whether the utterance presented is from the claimant or not.

### 3. Component Score Weighting

In classical GMM/UBM systems, both the hypothesized speaker and impostors are not well modeled due to insufficient data, and the mismatch between training and testing data is also one of the major reasons for detection errors. In this section, we will introduce our Component Score Weighting (CSW) approach to alleviate the two factors, in which, the log-likelihood score is first assigned to each Gaussian mixture component, and a nonlinear weighting function is applied to the component score serials.

#### 3.1 Component Score Representation

As shown in section 2, given a speaker model  $\lambda$  and a sequence of feature vectors  $X$ , the log-likelihood is the form of equation (2), while substitute  $p(\vec{x}_t|\lambda)$  with equation (1) giving

$$\log p(X|\lambda) = \frac{1}{T} \sum_t \log \left( \sum_{i=1}^M w_i p_i(\vec{x}_t) \right) \quad (5)$$

For each frame in the utterance, there exists one dominant component mixture, whose log-likelihood is maximal in all of the component mixture scores. For frame  $\vec{x}_t$ , suppose its corresponding dominant component mixture is indexed  $k$ , which satisfy  $w_k p_k(\vec{x}_t) \geq w_i p_i(\vec{x}_t), i = 1, \dots, M, i \neq k$ . then the log-likelihood of the frame  $\vec{x}_t$  against the speaker model  $\lambda$  can be rewritten as:

$$\frac{1}{T} \log p(\vec{x}_t|\lambda) = \frac{1}{T} \log \left( w_k p_k(\vec{x}_t) + \sum_{i=1, i \neq k}^M w_i p_i(\vec{x}_t) \right)$$

$$= \frac{1}{T} \log(w_k p_k(\vec{x}_t)) + \frac{1}{T} \log \left( 1 + \sum_{i=1, i \neq k}^M \frac{w_i p_i(\vec{x}_t)}{w_k p_k(\vec{x}_t)} \right) \quad (6)$$

$$= s_{kt} + \tilde{s}_{kt}$$

Where  $s_{kt} = \frac{1}{T} \log(w_k p_k(\vec{x}_t))$  denotes the dominant score of frame  $t$  which is dominated by the  $k_{th}$  Gaussian component mixture, while  $\tilde{s}_{kt} = \frac{1}{T} \log \left( 1 + \sum_{i=1, i \neq k}^M \frac{w_i p_i(\vec{x}_t)}{w_k p_k(\vec{x}_t)} \right)$  denotes the residual score of component  $k$  and frame  $t$ .

Sum all the frames together, we can define as:

$$s_k = \sum_{t=1}^T \delta_{t \in k} \frac{1}{T} \log(w_k p_k(\vec{x}_t)) \quad (7)$$

$$= \sum_{t=1}^T \delta_{t \in k} s_{kt}$$

and

$$\tilde{s}_k = \sum_{t=1}^T \delta_{t \in k} \frac{1}{T} \log \left( 1 + \sum_{i=1, i \neq k}^M \frac{w_i p_i(\vec{x}_t)}{w_k p_k(\vec{x}_t)} \right) \quad (8)$$

$$= \sum_{t=1}^T \delta_{t \in k} \tilde{s}_{kt}$$

where

$$\delta_{t \in k} = \begin{cases} 1 & \text{if the } k_{th} \text{ component is dominant for frame } t. \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

By this definition, we will get two component score vectors  $\vec{S}_d = [s_1, s_2, \dots, s_M]$  and  $\vec{S}_r = [\tilde{s}_1, \tilde{s}_2, \dots, \tilde{s}_M]$ , which are denoted as dominant component score vector and residual component score vector, respectively. And the log-likelihood of equation (5) can be rewritten as:

$$\log P(X|\lambda) = \sum_{k=1}^M \sum_{t=1}^T \delta_{t \in k} \frac{1}{T} \left( \log(w_k p_k(\vec{x}_t)) + \log \left( 1 + \sum_{i=1, i \neq k}^M \frac{w_i p_i(\vec{x}_t)}{w_k p_k(\vec{x}_t)} \right) \right) \quad (10)$$

$$= \sum_{k=1}^M (s_k + \tilde{s}_k)$$

#### 3.2 Extend The Original LLR By Weighting Function

Traditional log-likelihood score of the test segment and speaker model is a summation of the component scores, just as equation (10) shows. However, due to the insufficiency of the training data and the distortion of the test segment caused by channel, the Gaussian mixtures will have different discriminative capabilities, thus, the component scores do not have the same contribution to the final recognition performance, a simple summation of the those scores is not an optimum choice. In this paper, we use a

weighting function to reweight those scores, and extend the original log-likelihood as follows:

$$f(X|\lambda) = \sum_k^M (W_d(s_k)s_k + W_r(\tilde{s}_k)\tilde{s}_k) \quad (11)$$

where,  $W_d(\bullet)$  and  $W_r(\bullet)$  is the weighting function for the dominant component scores and the residual component scores, respectively. The weighting function is expected to reweight the component scores according to their discriminative capabilities, and for equation (10),  $W_d(\bullet)=1, W_r(\bullet)=1$ .

To estimate an appropriate weighting function in equation (11), we review the LLR detection function in equation (3), which can be rewritten as:

$$\begin{aligned} \Lambda(X) &= \log p(X|\lambda_{spk}) - \log p(X|\lambda_{ubm}) \\ &= \sum_{k=1}^M (s_k^{spk} + \tilde{s}_k^{spk}) - \sum_{k=1}^M (s_k^{ubm} + \tilde{s}_k^{ubm}) \end{aligned} \quad (12)$$

where  $s_k^{spk}$  and  $s_k^{ubm}$  indicate the dominant component scores of speaker model and UBM model, respectively (it's the same for the residues). To simplify the deduction, we consider the dominant scores only (the deduction process is similar for the residual scores):

$$\begin{aligned} \Lambda_d(X) &= \sum_{k=1}^M s_k^{spk} - \sum_{k=1}^M s_k^{ubm} \\ &= M \left( \sum x \frac{\#(s_k^{spk} = x)}{M} \right) - M \left( \sum x \frac{\#(s_k^{ubm} = x)}{M} \right) \\ &\approx M \int x p_{spk}(x) dx - M \int x p_{ubm}(x) dx \\ &= M \int x \{p_{spk}(x) - p_{ubm}(x)\} dx \\ &= M \int x \rho(x) dx \end{aligned} \quad (13)$$

where,  $\#(s_k = x)$  indicates the number of component scores whose value equal to  $x$ ,  $p_{spk}(x)$  is the Probability Density Function (PDF) of  $s_k^{spk}$ ,  $p_{ubm}(x)$  is the PDF of  $s_k^{ubm}$  and  $\rho(x) = p_{spk}(x) - p_{ubm}(x)$ .

The original LLR, just as equation (13) shows, is the integral of the product component score value  $x$  and the PDF difference  $\rho(x)$ . If we assume that the span of the component value  $x$  is the same for all the GMM models (the value of  $p(x)$  can be 0), then the value of the LLR will only depend on the PDF difference  $\rho(x)$ . When applied to the weighting functions in equation (11), a new estimation of the LLR derived from equation (13) will be as follows:

$$\begin{aligned} \Lambda'_d(X) &= \sum_k^M W_d(s_k^{spk})s_k^{spk} - \sum_k^M W_d(s_k^{ubm})s_k^{ubm} \\ &= M \sum W_d(x)x \left( \frac{\#(s_k^{spk} = x)}{M} - \frac{\#(s_k^{ubm} = x)}{M} \right) \\ &\approx M \int x W_d(x) \rho(x) dx \end{aligned} \quad (14)$$

Comparing with equation (13),  $\Lambda'_d(X)$  can be seen as an extended measure function of the original LLR, in which, the weighting function  $W_d(x)$  is added to reweight the PDF difference  $\rho(x)$ . If  $\rho(x)$  has the same discriminative capability over the span of  $x$ , then  $W_d(x)=1$  is optimal, and we will go back to the original LLR.

Thus, to estimate an appropriate weighting function  $W_d(x)$ , it's necessary to estimate the distribution of  $p_{spk}(x)$  and  $p_{ubm}(x)$  first, however, this is a difficult job, since many factors will affect the distribution of component scores, e.g. the number of Gaussian component mixtures  $M$  and the amount of training and testing data. In this paper, we investigate the weighting function based on the major task in speaker verification, in which, the training and testing data is relatively sufficient (about one minute or more) and the number of Gaussian mixtures in each GMM model is relatively large, e.g.  $M \in [512, 2048]$ . Limited to the length of the article, in this paper, we do not discuss the weighting function which is appropriate for sparse training and testing data, or small number of Gaussian mixtures. (Thus, the weighting function derived from the following section is not a universal one. A more general weighting function will be one of the focal points in our future work).

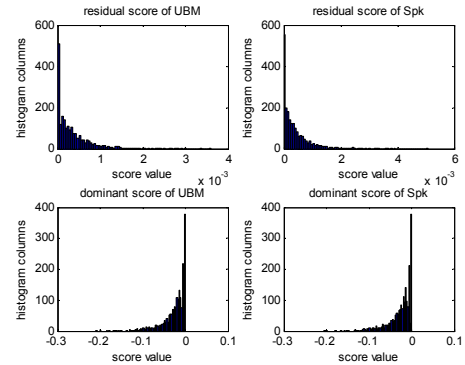


Figure 1: *histogram of dominant and residual score for UBM model and speaker model.*

### 3.3 Exponential Distribution Weighting

In this section, we investigate the weighting function appropriate for the major task in speaker verification, in which, the training and testing data is no less than 1 minute. Figure 1 indicates that the distribution of the component scores is approximately exponential when the mixture number  $M$  is 2048 in 1conv4w-1conv4w task of 2006 NIST SRE, in which, the training and testing data is about two minutes long. Consider the distribution

of the dominant component scores, we find most components get relatively high value, which perhaps due to the UBM model is well trained, and most Gaussian components in UBM can describe the testing data appropriately, thus the corresponding log-likelihood value is high. Similarly, it's straightforward that for a given speaker model, if the test utterance is presented by the hypothesis speaker, then more components will get relatively high log-likelihood value. There are two main reasons that will cause some of the components to get relatively low value, namely, those Gaussian components are not well trained or a mismatch exists between the training and testing data. Contrarily, if the test utterance comes from the impostors, then the number of high-valued component scores will be smaller, and the PDF of the component scores will tend to be smoother. Figure 2 illustrates the distribution of component scores in the two cases.

Based on the discussion above and the illustration of Figure 2, we find that the difference  $\rho(x)$  of dominant scores is more capable to distinguish the target speaker and the impostors in high-value interval of the span, and the low-valued component scores are more likely to express the variability caused by the channel or session difference, etc. Thus, the weighting function  $W_d(x)$  should be monotonously increasing, and according to the degree of weighting, different weighting functions can be used. In this paper, we simply used a normal exponential distribution as the weighting function, namely,  $W_d(x) = \exp(x)$  for the dominant scores, and investigated its performance. For the residues, the discussion will be more complicated. However, from experiments we find that the residual scores (weighted or not) have little discriminative capabilities, and empirically, we use  $W_r(x) = \exp(-x)$  as the weighting function, in this case, the sum of component scores after weighting will be as follows:

$$f(\tilde{S}_d) = \sum_{k=1}^M s_k \exp(s_k) \quad (15)$$

$$f(\tilde{S}_r) = \sum_{k=1}^M \tilde{s}_k \exp(-\tilde{s}_k) \quad (16)$$

After exponential weighting, equation (11) will be as follows:

$$f(X|\lambda) = f(\tilde{S}_d) + f(\tilde{S}_r) \quad (17)$$

However, the experiments show that, the dominant score is much more discriminative than the residual one, thus, we give a small scale factor on the residual score, and the final likelihood after weighting is as follows:

$$f(X|\lambda) = f(\tilde{S}_d) + \eta f(\tilde{S}_r), \quad \eta \in [0, 1] \quad (18)$$

Here  $\eta$  is the tunable smoothing weight which turns the proportion of dominant scores and residual scores. And the log-likelihood ratio of equation (3) can be rewritten as:

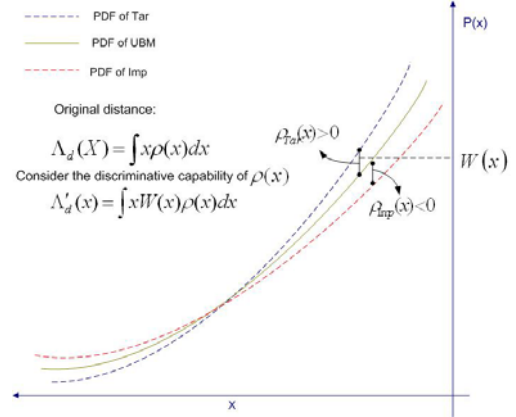


Figure 2: three kind of pdfs of the dominant component scores, namely, the pdf of target speakers, the pdf of impostors and the pdf of UBM model.

$$\begin{aligned} f(\Lambda) &= f_{spk}(\tilde{S}_d) + \eta f_{spk}(\tilde{S}_r) - f_{ubm}(\tilde{S}_d) - \eta f_{ubm}(\tilde{S}_r) \\ &= f_{spk}(\tilde{S}_d) - f_{ubm}(\tilde{S}_d) + \eta \left( f_{spk}(\tilde{S}_r) - f_{ubm}(\tilde{S}_r) \right) \end{aligned} \quad (19)$$

## 4. Experimental Results

In this section, we will report experimental results on GMM based speaker verification system using Component Score Weighting (CSW). Section 4.1 presents the datasets used in our experiments. The result of these experiments is discussed in section 4.2.

### 4.1 Datasets Description

For cepstral feature extraction, 13-dimensional PLP vectors were calculated from the silence removed speech signal every 10ms using a 25ms Hamming window. Band-limiting was performed by only retaining the filterbank outputs from the frequency range 300Hz-3400Hz. Cepstral features were processed with RASTA filtering to eliminate channel distortion. Delta, acceleration and triple-delta coefficients were then computed over  $\pm 2$  frames span and appended to each feature vector, which results in a dimensionality of 52. Feature mapping and histogram equalization (HEQ) were applied to improve channel and noise robustness. Heteroscedastic linear discriminant analysis (HLDA) was then used to decorrelate the features and reduce the dimensionality from 52 to 51(1 dimension is left out as nuisance).

Speaker verification experiments were conducted on the 2006 NIST SRE corpus [10]. We focus on male part of the single-side 1 conversation train, single-side 1 conversation task, which contains 1570 true trails and 20561 false trails. A gender independent UBM with 2048 Gaussians was used in all the experiments, which was trained using about 40 hours of data from the Switchboard II corpora (phase 1&2). The speaker GMM models were adapted from UBM by MAP adaptation with the relevance fact set to be 16 (only the means were adapted).

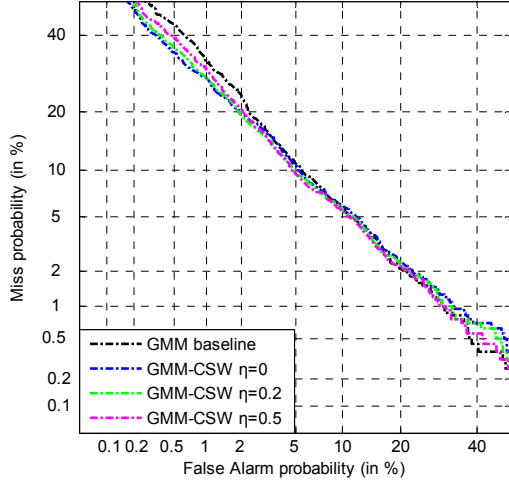


Figure 3: DET curves for GMM baseline and GMM with CSW in the 1conv4w-1conv4w task of the 2006 NIST SRE without TNorm.

Results are presented using Detection Error Tradeoff (DET) plots. Along with Equal Error Rate (EER), the minimum detection cost function (DCF) value, as defined by NIST [10], is also used as an overall performance measure.

#### 4.2 Results And Discussion

Figure 3 shows the DET curves of GMM baseline system and GMM with Component Score Weighting (CSW) in 1conv4w-1conv4w task of the 2006 NIST SRE. We compared the performance of GMM-CSM with the dominant score only ( $\eta=0$  in equation (18)), and GMM-CSM with both the dominant and residual score. Table 1 gives the results of the three systems in terms of both minimum detection cost function (DCF) and equal error rate (EER). DCF is the Bayesian risk function defined by NIST with  $P_{t, \text{arg et}} = 0.01$ ,  $C_{fa} = 1$  and  $C_{miss} = 10$  [10].

Table1. Results for GMM baseline and GMM with Component Score Weighting without TNorm

system	EER (%)	MinDCF (x100)
GMM baseline	7.64	4.16
GMM with CSW $\eta = 0$	<b>7.45</b>	<b>3.66</b>
GMM with CSW $\eta = 0.2$	7.25	3.71
GMM with CSW $\eta = 0.5$	7.26	3.84

Table2. Results for GMM baseline and GMM with Component Score Weighting with TNorm

system	EER (%)	MinDCF (x100)
GMM baseline	7.64	4.16
GMM with CSW	7.45	3.66
GMM with TNorm	6.96	3.48
GMM with CSW&TNorm	<b>7.14</b>	<b>3.10</b>

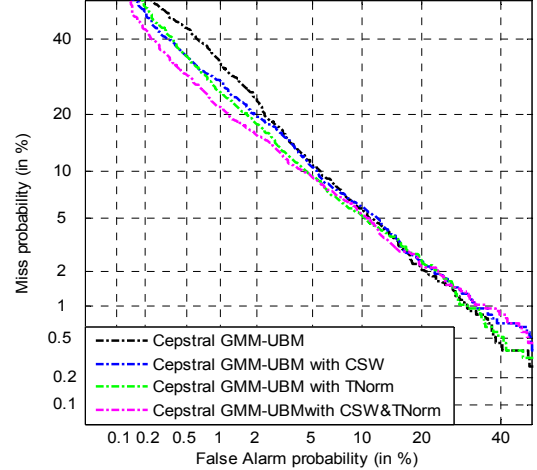


Figure 4: DET curves for GMM baseline and GMM with CSW in the 1conv4w-1conv4w task of the 2006 NIST SRE with TNorm.

From Table 1, we can find that, when applied Component Score Weighting (CSW), obvious system performance gains will be achieved comparing with the GMM/UBM baseline system, especially in minimum DCF. And when using the dominant score only to perform the verification, we get the best system performance (about 12% relative improvement in minimum DCF), however, when adding the residual score, the system performance will be degraded, although slightly better in EER.

This result perhaps is due to the effect of Component Score Weighting (CSW). The dominant score serials derived from equation (7) are reasonable, however, assigning the residual scores to each Gaussian mixture according to equation (8) is not the same case. Thus, the exponential weighting function is only useful to the dominant scores, but not the residual one. A better representation of the residual scores is needed, however, just as the experiment shows, discarding the residual scores will not affect the system performance.

To further examine the performance of Component Score Weighting (CSW), we performed a set of experiments with TNorm. Considering the low discriminative capabilities of the residual scores, we discarded them and use the dominant scores only in these experiments. Figure 4 shows the DET curves of GMM baseline and GMM-CSW with TNorm. Table 2 gives the results in terms of EER and the minimum DCF. Experiments show that, Component Score Weighting (CSW) can still improve the system performance obviously after TNorm, and reduce the minimum DCF from  $34.8 \times 10^{-3}$  to  $31.0 \times 10^{-3}$  (about 11% relative improvement). That means the exponential distribution weighting function is appropriate in this situation.

## 5. Conclusion and Future works

This paper proposed a novel Component Score Weighting (CSW) approach for state-of-the-art GMM/UBM speaker verification. The log-likelihood score of the test utterance against speaker model was first assigned to each Gaussian component mixture, deriving two component score serials, and then a weighting function was applied to reweight those scores according to their discriminative capability. This paper also discussed the appropriate weighting function and proposed the exponential weighting approach when the training and testing

data was relatively sufficient (about one minute or more). Experiments on NIST 2006 corpus show that, our approach achieves relative improvements of up to 12% in minimum decision cost function (DCF) over our previous GMM/UBM baseline.

Just as discussed in the paper, the exponential weighting function is not a universal one, in some tasks, e.g. the training and testing data is sparse, the weighting function will not be appropriate. Thus, in the future, our work will focus on a more general weighting function which can be adopted in different situations. Additionally, the channel and session variability can be alleviated in component score space, and future work in this direction will also be proposed.

## 6. References

- [1] D. Reynolds, T. Quatieri and R. Dunn, "Speaker verification using adapted Gaussian Mixture Models," *Digital Signal Processing*, vol.10, pp. 19-41, 2000.
- [2] V. Wan and S. Renals, "Speaker verification using sequential discriminant support vector machines," *IEEE Trans. Speech and Audio Processing*, vol. 13, no. 2, pp. 203-210, 2005.
- [3] W. M. Campbell, "Generalized linear discriminant sequence kernels for speaker recognition," in *Proc. ICASSP*, 2002, pp. 161-164.
- [4] M. Campbell, D. E. Sturim, D. A. Reynolds and A. Solomonoff, "SVM-based speaker verification using a GMM supervector kernel and NAP variability compensation," in *Proc. ICASSP*, 2006, pp. 97-100.
- [5] H. Hermansky, N. Morgan, "RASTA processing of speech," *IEEE Trans. on Speech and Audio Processing*, vol.2, no.4, pp. 578~589, Oct. 1994.
- [6] D. Reynolds, "Channel robust speaker verification via feature mapping," in *Proc. ICASSP*, 2003, pp. 53-56.
- [7] Heck, L., and Weintraub, M, "Handset dependent background models for robust text-independent speaker recognition," In *Proc. ICASSP 1997, Munich, 1997, Vol. 2*, pp. 1071-1074.
- [8] R. Auckenthaler, M.Carey, and H. Lloyd-Thomas, "Score Normalization for Text-Independent Speaker Verification Systems, *Digital Signal Processing*, vol. 10, pp. 42-54, January 2000.
- [9] D. E. Sturim and D. A. Reynolds, "Speaker adaptive cohort selection for Tnorm in text-independent speaker verification," In *Proc. ICASSP*, 2005, pp. 741-744.
- [10] "The NIST 2006 speaker recognition evaluation plan," <http://www.nist.gov/speech/tests/spk/spk/2006/>