Benefit of Maximum Likelihood Linear Transform (MLLT) Used at Different Levels of Covariance Matrices Clustering in ASR Systems[*]

Josef V. Psutka

University of West Bohemia, Faculty of Applied Sciences,
Department of Cybernetics,
Univerzitní 8, 306 14 Plzeň, Czech Republic
psutka_j@kky.zcu.cz

Abstract. The paper discusses the benefit of a Maximum Likelihood Linear Transform (MLLT) applied on selected groups of covariance matrices. The matrices were chosen and clustered using phonetic knowledge. Results of experiments are compared with outcomes obtained for diagonal and full covariance matrices of a baseline system and also for widely used transforms based on Linear Discriminant Analysis (LDA), Heteroscedastic LDA (HLDA) and Smoothed HLDA (SHLDA).

1 Introduction

The absolute majority of current LVCSR systems work with acoustic models based on hidden Markov Models (HMMs). Output distributions tied to the states of the model and expressed by multidimensional Gaussian distributions (simply by "Gaussians") or, more exactly, by the mixtures of Gaussians are considered to be a fundamental attribute of this concept. The application of a mixture of Gaussians for modeling an output distribution results from an effort to both catch the possible non-Gaussian nature of density functions which are associated with a particular state and model mutual correlations of elements in feature vectors.

Since ideally all output distributions should be computed for each incoming feature vector, it is useful notably for real-time applications to reduce the huge amount of computations which increase with a size of the dimension of a feature space and also with the number of Gaussians. To reduce the computation burden associated with evaluating output distributions, we can apply some of the following techniques:

- To execute decorrelation of feature vectors and to use diagonal rather a full covariance matrices (CMs) for the modeling of output distributions. For these purposes, usually some orthogonal transform based on the DCT (Discrete Cosine Transform) or the NPS (Normalization of Pattern Space) is applied [1].

[*] This paper was supported by the AVCR, project no. 1QS101470516., GACR, project no. 102/05/0278 and the project of the EU 6^{th} FP no. IST-034434.

V. Matoušek and P. Mautner (Eds.): TSD 2007, LNAI 4629, pp. 431–438, 2007.

- To reduce the dimension of pattern space using the projection of feature vectors from the original space to the space with lower dimension. A typical approach is based on PCA (Principal Component Analysis), LDA (Linear Discriminant Analysis) or HLDA (Heteroscedastic LDA).
- To change over from the triphone- to the monophone-based structure of HMMs where an influence of suppressed dependencies among features is alleviated mainly by enhancing the number of Gaussians in the individual states of monophone models. LVCSR systems with a triphone-based structure work typically with 30 up to 120 thousand Gaussians, whereas systems working with monophone-based structure use from 5 to 15 thousand Gaussians [2].

Naturally, there are many other clever approaches which speed up computations or choose only relevant states with associated Gaussians for evaluations. Generally, it is possible to say that both the pass from the triphone- to the monophone-based concept on the one hand and the various transformation techniques on the other hand decrease the number of computations, but they simultaneously bring about increasing the word error rate (WER) in comparison with using full CMs. Moreover, in case of transforms applied in a level of feature vectors, it is usually unfeasible to find the only one transformation which could decorrelate all elements of feature vectors of all states.

Recently, new approaches have been designed which alleviate the above-mentioned increasing the WER, while simultaneously preserving a relatively high computation efficiency. These techniques, known as Maximum Likelihood Linear Transform (MLLT) [3], [4] and Semi-Tied Covariance (STC) [7], suppose one transformation matrix to be tied with a group of covariance matrices belonging to (an) individual state(s) of a HMM (i.e. a set of transform matrices is used for the accomplishment of the transformation).

This paper describes experiments which were performed using the MLLT applied on selected groups of CMs (i.e. multiple application of a MLLT principle), which should be decorrelated. The covariance matrices were chosen and clustered using phonetic knowledge. The results of experiments are compared with outcomes obtained for diagonal and full covariance matrices and also for transforms based on LDA, HLDA and SHLDA (Smoothed HLDA) [5].

2 Maximum Likelihood Linear Transform (MLLT)

Current ASR systems use HMMs with continuous parameters which are represented for each state by a Gaussian Mixture Model (GMM). A standard GMM with parameters given by $\Theta = \{c_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\}_{j=1}^{M}$ it is of the form

$$p(\boldsymbol{x}|\Theta) = \sum_{j=1}^{M} c_j \; N(\boldsymbol{x}; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j), \tag{1}$$

where M is the number of components in a mixture, c_j is the j-th component weight satisfying requirements: $c_j \geq 0$, $\sum_{j=1}^{M} c_j = 1$, $\boldsymbol{\mu}_j$ is the mean value of the j-th component and $\boldsymbol{\Sigma}_j$ is a square covariance matrix of the j-th component of the rank n. As was mentioned in Section 1, almost all currently used systems work with the CM of a diagonal form. In comparison with a full covariance concept, this approach has an evident advantage especially owing to lower computational and storage burden and also due to a robust parameter estimation. However, the diagonal concept of CMs can be fully beneficial only on condition that elements of the feature vector are mutually independent. The MLLT introduces a new form of a CM which allows sharing a few full covariance matrices over many distributions. Instead of having a distinct CM for every component in the recognizer, each CM consists of two elements: a non-singular linear transformation matrix $\boldsymbol{W}^{\mathrm{T}}$ shared over a set of components, and the diagonal elements in the matrix $\boldsymbol{\Lambda}_j$. The inverse covariance (precision) matrix $\boldsymbol{\Sigma}_j^{-1}$ is then of the form

$$\boldsymbol{\Sigma}_j^{-1} \approx \boldsymbol{W}\boldsymbol{\Lambda}_j\boldsymbol{W}^{\mathrm{T}} = \sum_{k=1}^{n} \lambda_j^k \boldsymbol{w}_k \boldsymbol{w}_k^{\mathrm{T}}, \tag{2}$$

where $\boldsymbol{\Lambda}_j$ is a diagonal matrix with entries $\boldsymbol{\Lambda}_j = \mathrm{diag}(\boldsymbol{\lambda}_j) = \mathrm{diag}(\lambda_j^1, \lambda_j^1, \ldots, \lambda_j^n)$ and $\boldsymbol{w}_k^{\mathrm{T}}$ is the k^{th} row of the transformation matrix $\boldsymbol{W}^{\mathrm{T}}$. Estimations of model parameters could be performed by the maximum likelihood approach. We look for such a set of parameters $\hat{\Theta}$, which satisfies the equation

$$\hat{\Theta} = \underset{\Theta}{\mathrm{argmax}} \sum_{i=1}^{N} \log\ p(\boldsymbol{x}_i|\Theta). \tag{3}$$

The solution of this equation cannot be determined in an analytical form. However, we can use an iterative procedure based on the EM algorithm.

2.1 Levels of CMs Clustering

The MLLT (or perhaps better the multiple-MLLT) proposed in this article supposes the transform matrix $\boldsymbol{W}^{\mathrm{T}}$ to be generally searched for each of R selected groups of CMs. For a better insight into this problem Fig. 1 illustrates examples of using transformation matrices $\boldsymbol{W}_r^{\mathrm{T}}$ in a HMM structure of an ASR system. A diagram shows five levels of a tree structure based on a successive division of a whole set of GMMs. In fact, this division is based on phonetic knowledge and supposes the 3-state HMM of a triphone to be used as a basic unit for acoustic modeling.

It is evident that the whole acoustic model ”•−•+•” consists of a set of particular triphone models ”•−?+•”. In a simplified way all triphones with the same centre phoneme can be considered to be a generalized model of a monophone, e.g. the phoneme ”A” depicted in the form ”•−A+•”. Considering different frequency of occurrence of individual triphones in training data it is pertinent owing to a robust estimation of statistics of less frequent units to cluster phonetically similar states of models (usually using a phonetic decision tree). It means that each state of one generalized monophone can include several clustered representatives each of them is described by an individual mixture (GMM) equipped with M components. Individual components of a GMM can be described by probability density functions of the form $N(\boldsymbol{x}; \boldsymbol{\mu}_{\Phi(\alpha)\_\beta\_\gamma}, \boldsymbol{\Sigma}_{\Phi(\alpha)\_\beta\_\gamma})$, where $\boldsymbol{x}$ is the feature vector, $\boldsymbol{\mu}_{\Phi(\alpha)\_\beta\_\gamma}$, is the mean value and $\boldsymbol{\Sigma}_{\Phi(\alpha)\_\beta\_\gamma}$ is the covariance matrix of the γ-th component of the β-th representative (mixture) of the state α of the phoneme Φ.

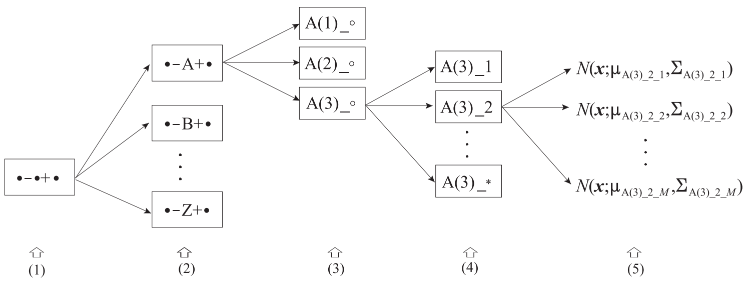
•-•+•
•-A+•
•-B+•
•-Z+•
A(1)_∘
A(2)_∘
A(3)_∘
A(3)_1
A(3)_2
A(3)_*
N(x;μ_A(3)_2_1,Σ_A(3)_2_1)
N(x;μ_A(3)_2_2,Σ_A(3)_2_2)
N(x;μ_A(3)_2_M,Σ_A(3)_2_M)
(1)
(2)
(3)
(4)
(5)

Fig. 1. Levels of CMs clustering

As was described above there are basically five different levels of CMs clustering, which are indicated in Fig. 1:

- $MLLT(1)$: There is only one transformation matrix for all components of all GMMs (mixtures) of all states of all generalized monophones.
- $MLLT(2)$: There is a separate transformation matrix for all components of all mixtures of all states of an individual generalized monophone (it means that each generalized monophopne has its own transformation matrix).
- $MLLT(3)$: In this case, a separate transformation matrix is connected to all components of all mixtures belonging to an individual state of a given generalized monophone.
- $MLLT(4)$: There is a separate transformation matrix for all components of a given mixture which belongs to an individual state of a given generalized monophone.
- $MLLT(5)$: Each component in all mixtures of HMMs has its own transformation matrix. This approach is equivalent to the case of GMMs equipped with the full CMs, because each symmetric positive definite matrix has clear decomposition to the diagonal and transformation matrices (eigen values and eigen vectors).

2.2 Practical Application of the MLLT for Different Levels of CMs Clustering

Let us suppose R groups of CMs to be selected in a given level of clustering. It means that each group of CMs will share its own transformation matrix $\boldsymbol{W}_r^{\mathrm{T}}$, $r = 1, \ldots, R$. As was mentioned above, to estimate parameters of HMMs including parameters of transformation matrices $\boldsymbol{W}_r^{\mathrm{T}}$, $r = 1, \ldots, R$, we will have to use the EM algorithm. This algorithm supposes the construction of an optimization function $Q(\bar{\boldsymbol{\Theta}}, \boldsymbol{\Theta})$, which can be expressed in the form [6]

$$Q(\bar{\boldsymbol{\Theta}}, \boldsymbol{\Theta}) = \sum_{r=1}^{R} \sum_{(s,j) \in \Xi^{(r)}} \sum_{t} \gamma_{sj}(t) \log \left[c_{sj}\ N(\boldsymbol{o}(t); \boldsymbol{\mu}_{sj}, \boldsymbol{\Lambda}_{sj}, \boldsymbol{W}_r^{\mathrm{T}}) \right] =$$

$$= \sum_{r=1}^{R} \sum_{(s,j) \in \Xi^{(r)}} \sum_{t} \gamma_{sj}(t) \{ \log c_{sj} - \frac{n}{2} \log(2\pi) + \frac{1}{2} \log \det \left(\boldsymbol{W}_r\ \boldsymbol{\Lambda}_{sj}\ \boldsymbol{W}_r^{\mathrm{T}} \right) - \quad (4)$$

$$- \frac{1}{2} \left(\boldsymbol{o}(t) - \boldsymbol{\mu}_{sj} \right)^{\mathrm{T}}\ \boldsymbol{W}_r\ \boldsymbol{\Lambda}_{sj}\ \boldsymbol{W}_r^{\mathrm{T}}\ \left(\boldsymbol{o}(t) - \boldsymbol{\mu}_{sj} \right) \},$$

where $\bar{\boldsymbol{\Theta}} = \{\bar{\Theta}_r\}_{r=1}^{R}$ is a set of parameters of the old and $\boldsymbol{\Theta} = \{\Theta_r\}_{r=1}^{R}$ parameters of the new system. Let us remark that $\Theta_r = \left\{c_{sj}, \boldsymbol{\mu}_{sj}, \boldsymbol{\Lambda}_{sj}, \boldsymbol{W}_r^{\mathrm{T}}\right\}_{(s,j)\in\Xi^{(r)}}$ have their own transformation matrix $\boldsymbol{W}_r^{\mathrm{T}}$. $\Xi^{(r)}$ is a set of couples (s, j), where s corresponds to the state and j is an index of a component. In fact s matches the state of an original triphone model which is, in a sense of above described clustering process, consecutively clustered in various ways (Note: Because each state of an original triphone model was always represented by one mixture (GMM), we can alternatively see the index s as a label of a concrete mixture in the system of all mixtures representing the whole HMM of a task); $\gamma_{sj}(t)$ is the a posteriori probability of the j-th component and the s-th state given an observation of the feature vector $\boldsymbol{o}(t)$ and the set of parameters $\boldsymbol{\Theta}$; $\boldsymbol{\Lambda}_{sj}$ is the diagonal matrix, $\boldsymbol{\mu}_{sj}$ is the mean value and c_{sj} is the a priori probability of the s-th state and the j-th component.

Let us have a look at the set $\Xi^{(r)}$ from the point of view of the variant $MLLT(2)$. In this case the R would be equal to the number of phonemes (in the phoneme alphabet) and one set $\Xi^{(r)}$ would consist of all components of all states which are connected with all triphones containing the same centre phoneme.

Searching for maximum of $Q(\bar{\boldsymbol{\Theta}}, \boldsymbol{\Theta})$ respecting c_{sj}, $\boldsymbol{\mu}_{sj}$ and $\boldsymbol{\Sigma}_{sj}$ results in expected equations

$$\hat{c}_{sj} = [\sum_t \gamma_{sj}(t)]\ [\sum_j \sum_t \gamma_{sj}(t)]^{-1} \tag{5}$$

$$\hat{\boldsymbol{\mu}}_{sj} = [\sum_t \gamma_{sj}(t)\boldsymbol{o}(t)]\ [\sum_t \gamma_{sj}(t)]^{-1} \tag{6}$$

$$\hat{\boldsymbol{\Sigma}}_{sj} = [\sum_t \gamma_{sj}(t)(\boldsymbol{o}(t) - \hat{\boldsymbol{\mu}}_{sj})\ (\boldsymbol{o}(t) - \hat{\boldsymbol{\mu}}_{sj})^{\mathrm{T}}]\ [\sum_t \gamma_{sj}(t)]^{-1} \tag{7}$$

Maximization of (4) simultaneously with respect to $\boldsymbol{\Lambda}_{sj}$ and $\boldsymbol{W}_r^{\mathrm{T}}$ is possible only in quite trivial cases. For that reason, the maximum has to be estimated by a special iterative procedure [6], [7]. This technique consists firstly in optimization of $\boldsymbol{\Lambda}_{sj}$ at fixed $\boldsymbol{W}_r^{\mathrm{T}}$ and in the next step, on the contrary, the $\boldsymbol{W}_r^{\mathrm{T}}$ is optimized at the fixed $\boldsymbol{\Lambda}_{sj}$, which results in the estimate

$$\hat{\boldsymbol{\Lambda}}_{sj} = \left[\mathrm{diag}\ (\boldsymbol{W}_r^{\mathrm{T}} \hat{\Sigma}_{sj} \boldsymbol{W}_r)\right]^{-1}, \tag{8}$$

where $(s, j) \in \Xi^{(r)}$, $\hat{\boldsymbol{\Sigma}}_{sj}$ is the estimate of a covariance matrix of the j-th component and the s-th state. An iterative procedure for an estimate of an individual row of the transformation matrix $\boldsymbol{W}_r^{\mathrm{T}}$ can be written in the form of

$$(\tilde{\boldsymbol{w}}_k^{(r)})^{\mathrm{T}} = (c_k^{(r)})^{\mathrm{T}}\ (\boldsymbol{G}_k^{(r)})^{-1} \sqrt{\frac{T}{(c_k^{(r)})^{\mathrm{T}} (\boldsymbol{G}_k^{(r)})^{-1} (c_k^{(r)})}}, \tag{9}$$

where

$$\boldsymbol{G}_k^{(r)} = \sum_{(s,j)\in\Xi^{(r)}} \sum_t \hat{\lambda}_{sj}^k \gamma_{sj}(t) \left[\boldsymbol{o}(t) - \hat{\boldsymbol{\mu}}_{sj}\right]\ \left[\boldsymbol{o}(t) - \hat{\boldsymbol{\mu}}_{sj}\right]^{\mathrm{T}}, \tag{10}$$

$k = 1, \ldots, n$; n is a dimension of the feature space, T is a number of speech feature vectors, $\hat{\lambda}_{sj}^k$ is the k-th diagonal element of the matrix $\hat{\boldsymbol{\Lambda}}_{sj}$, $(\tilde{\boldsymbol{w}}_k^{(r)})^{\mathrm{T}}$ is the estimate of the k-th row of the transformation matrix $\boldsymbol{W}_r^{\mathrm{T}}$, $(c_k^{(r)})^{\mathrm{T}}$ is the k-th row of the cofactor matrix to the transformation matrix $\boldsymbol{W}_r^{\mathrm{T}}$, $\gamma_{sj}(t)$ is the a posteriori probability of the j-th component and the s-th state given an observation of a feature vector $\boldsymbol{o}(t)$ at the time t and a set of parameters $\boldsymbol{\Theta}$. Let us describe one pass through the modified EM algorithm:

1. Using up-to-date set of parameters we compute a posteriori probabilities $\gamma_{sj}(t)$ for all components j, all states s and all times t (this point corresponds to a standard step of the Baum-Welch procedure).
2. On the basis of $\gamma_{sj}(t)$ we estimate for all j and s a priori probabilities $\hat{c}_{sj}$ (5), mean values $\hat{\boldsymbol{\mu}}_{sj}$ (6) and covariance matrices $\hat{\boldsymbol{\Sigma}}_{sj}$, see (7).
3. Using current transformation matrices $\boldsymbol{W}_r^{\mathrm{T}}$ we compute new estimates of all diagonal matrices $\hat{\boldsymbol{\Lambda}}_{sj}$ according to (8).
4. We estimate all transformation matrices $\boldsymbol{W}_r^{\mathrm{T}}, r = 1, \ldots, R$. To accomplish this step, a special iterative procedure must be run. For individual transformation matrices we compute:
   - the matrices $\boldsymbol{G}_k^{(r)}$ according to (10), where $k = 1, \ldots, n$ and n is the dimension of a feature space,
   - the cofactor matrix $\boldsymbol{W}_r^{\mathrm{COF}}$ to the current estimate of the transformation matrix $\boldsymbol{W}_r^{\mathrm{T}}$,
   - the individual rows of the transformation matrix $\tilde{\boldsymbol{W}}_r^{\mathrm{T}}$ according to (9).

   Then we update $\boldsymbol{W}_r^{\mathrm{T}}$ and continue the iterative procedure from the point of 4a) until the convergence is reached.
5. We replace the old estimate of the parameters $\bar{\boldsymbol{\Theta}}$ by the new $\boldsymbol{\Theta}$ ones and repeat by the step 1 until the convergence of the EM algorithm is reached.

As soon as the transform matrices $\boldsymbol{W}_r^{\mathrm{T}}$, $r = 1, \ldots, R$, are estimated, we can use them for transformation of feature vectors $\boldsymbol{o}(t)$ to the new feature space. Generally, the feature vector $\boldsymbol{o}(t)$ should be put as many transformations

$$\boldsymbol{o}(t)^{(r)}(t) = \boldsymbol{W}_r^{\mathrm{T}}\, \boldsymbol{o}(t), \quad \text{for } r = 1, \ldots, R, \tag{11}$$

as many groups R of CMs are in a given task clustered. The resulting likelihood function $\log N(\boldsymbol{o}(t); \boldsymbol{W}_r^{\mathrm{T}}\boldsymbol{\mu}_{sj}, \boldsymbol{\Lambda}_{sj}^{-1})$ must then be normalized by a likelihood compensation term, which is proportional to $\log \det \boldsymbol{W}_r^{\mathrm{T}}$.

3 Results of Experiments

All the experiments were performed using a speech data set of telephone quality. The corpus consists of Czech read speech transmitted over a telephone channel. One thousand speakers were asked to read various sets of 40 sentences. The digitization of an input analog telephone signal was provided by a telephone interface board DIALOGIC D/21D at 8 kHz sample rate and converted to the mu-law 8 bit resolution. The telephone test set consisted of 100 sentences randomly selected from utterances of 100 different

Table 1. Results of comparative experiments

	# of transf. matrices	WER[%]	#of estimated parameters
DIAG	-	14.03	≈2.33M
LDA dim 26	1	13.74	≈1.68M
HLDA dim 25	1	13.81	≈1.62M
SHLDA dim 25	1	13.96	≈1.62M
MLLT(1)	1	11.68	≈2.33M
MLLT(2)	43	11.64	≈2.38M
MLLT(3)	129	11.46	≈2.50M
MLLT(4)	4044	11.29	≈7.57M
FULL	-	9.18	≈22.70M

speakers who were not included in the telephone training database. The lexicon in all test tasks contained 475 different words. Since several words had multiple different phonetic transcriptions, the final vocabulary consisted of 528 items. There were no OOV words.

The front-end of the ASR system is based on the MFCC parameterization. Feature vectors consisted of 12 static + 12 delta + 12 delta-delta = 36 coefficients (including the zeroth cepstral coefficient representing the signal energy) were computed with a frame spacing of 10ms. The number of coefficients in a feature vector and the number of band-pass filters (f =15) applied in the frequency axis (0÷4kHz) is a result of a thorough analysis and extensive experimental works in which robust setting of the MFCC-based parameterization was searched for [6]. Cepstral mean normalization (CMN) was used to reduce the effect of constant channel characteristics. No variance or vocal tract length normalization (VTLN) was applied in these experiments.

The basic speech unit in all the experiments was a triphone. Each individual triphone is represented by a 3-state HMM; each state is provided by a mixture of 8 components of a multivariate Gaussians. In all recognition experiments a language model based on zero-grams was applied so that the influence of individual transforms could be better judged. For that reason, the perplexity of the task was 528.

The goal of following experiments was to explore how an application of different levels of CMs clustering influences recognition results (WER). In order to judge a benefit of the Maximum Likelihood Linear Transform used at different levels of CMs clustering, we also performed comparative experiments on a standard baseline ASR system with CMs of the diagonal form (DIAG) and with the system working with the full CMs (FULL). In addition, several comparative tests were also made with widely used linear transforms based on LDA, HLDA and SHLDA. In these experiments we searched for the lowest dimension of the feature space which yielded lower or equal WER than the solution based on DIAG. In the case of SHLDA we set the smoothing factor α on a recommended value, which is $\alpha = 0.8$ [5]. The Results of all the experiments are itemized in Table 1 together with information about the number of parameters that have to be estimated.

4 Conclusions

The Maximum Likelihood Linear Transform (MLLT) applied at different levels of covariance matrices clustering (i.e. multiple applications) is a very effective technique of parameter decorrelation, which overcomes the widely used Linear Discriminant Analysis (LDA), Heteroscedastic LDA (HLDA) and also SHLDA (Smoothed HLDA) and approaches the application with the full CMs. However, results shown in Table 1 indicate that the growth of a number of transform matrices also causes the growth of a number of parameters that should be estimated, as well as the increase in the number of computations, especially during enumeration of output distributions. Therefore, the tradeoff between computational burdens (e.g. a case of real time applications) and the WER is necessary. We can also consider a different concept of CMs clustering which would not be based on phonetic knowledge but rather directly on real data, i.e. the ”geometrical” form of covariance matrices. This concept could bring the same or better results simultaneously with decreasing the number of transformation matrices.

References

1. Psutka, J.V., Müller, L.: Comparison of various decorrelation techniques in automatic speech recognition. Jour. of Syst., Cyb. and Inf. 5(1), 27–30 (2007)
2. Psutka, J.V., Müller, L.: Building Robust PLP-based Acoustic Module for ASR Application. In: Proc. of the 10^{th} SPECOM’2005, Greece, pp.761–764 (2005)
3. Olsen, P.A., Gopinath, R.A.: Modeling inverse covariance matrices by basis expansion. IEEE Trans. on Speech and Audio Proc. 12(1), 272–281 (2004)
4. Visweswariah, K., Axelrod, S., Gopinath, R.A.: Acoustic modeling with mixtures of subspace constrained exponential models. In: Proc. of the 7^{th} Eurospeech’2003, Geneva, Switzerland, pp. 2613–2616 (2003)
5. Burget, L.: Combination of Speech Features Using Smoothed Heteroscedastic Linear Discriminant Analysis. In: Proc. of the 8^{th} ICSLP’2004, Jeju, Korea, pp. 2549–2552 (2004)
6. Psutka, J.V.: Techniques of parameterization, decorrelation and dimension reduction in ASR systems. Thesis. Depart. of Cybernetics, Univ. of West Bohemia, Pilsen (in Czech) (2007)
7. Gales, M.J.F.: Semi-Tied Covariance Matrices for Hidden Markov Models. IEEE Transactions on Speech and Audio Processing 7(3), 272–281 (1999)