FEATURE TRANSFORMATION BASED ON DISCRIMINANT ANALYSIS PRESERVING LOCAL STRUCTURE FOR SPEECH RECOGNITION

Makoto Sakai[1,2]*, Norihide Kitaoka*[2]*, Kazuya Takeda*[2]

[1] DENSO CORPORATION, Nisshin 470-0111, Japan
[2] Nagoya University, Nagoya 464-8601, Japan

ABSTRACT

To improve speech recognition performance, a feature transformation based on discriminant analysis has been widely used to reduce redundant dimensions of features. Linear discriminant analysis (LDA) and Heteroscedastic discriminant analysis (HDA) are often used for this purpose, and a generalization method for LDA and HDA called Power LDA (PLDA) has been proposed. However, these methods may result in unexpected dimensionality reduction for multimodal data. It is important to preserve the local structure of the data in reducing the dimensionality of multimodal data. In this paper we introduce two methods, locality preserving HDA and locality preserving PLDA. We also give an efficient calculation scheme to obtain an optimal projection.

Index Terms— Speech recognition, Feature extraction, Multidimensional signal processing

1. INTRODUCTION

Hidden Markov Models (HMMs) have been widely used to model speech signals for speech recognition. However, they cannot precisely model the time dependency of features. In order to overcome this limitation, several researchers have proposed extensions, e.g., segmental unit input HMM [1]. In segmental unit input HMM, the immediate use of several successive frames as an input vector inevitably increases the number of dimensions. The concatenated vectors often have strong correlations among dimensions, and often include nonessential information. In addition, high-dimensional data require a heavy computational load. Therefore, to reduce dimensionality, a feature transformation method is often applied.

Linear discriminant analysis (LDA), also known as Fisher discriminant analysis (FDA), is widely used to reduce dimensionality and is a powerful tool to preserve discriminative information [2, 3]. In the speech recognition community, heteroscedastic discriminant analysis (HDA) is also used to reduce dimensionality [4]. HDA employs individual weighted contributions of the classes for its objective function. In addition, a generalization method for LDA and HDA has been proposed, which is called power LDA (PLDA) [5].

These methods may result in unexpected dimensionality reduction if the data in a certain class consist of several clusters, i.e., multimodality. In speech recognition, speech signals for acoustic model training tend to be multimodal because they are generally collected under various conditions, such as gender, age and noise environment. Therefore, a dimensionality reduction method for multimodal data is desired for improving speech recognition performance.

Recently, several methods have been proposed to reduce dimensionality of multimodal data in the machine learning community [6, 7]. It is important to preserve the local structure of the data in reducing the dimensionality of multimodal data appropriately. Locality preserving projection (LPP) [6] finds a projection such that the data pairs close to each other in the original space remain close in the projected space. Thus, LPP reduces dimensionality without losing information on local structure. Local Fisher discriminant analysis (LFDA) [7] is also proposed as a supervised method for multimodal data, while LPP is an unsupervised method. LFDA combines the ideas of FDA and LPP and maximizes between-class separability and preserves within-class local structure.

In this paper, inspired by LFDA, we combine the ideas of LPP and HDA. In addition, we also combine the ideas of LPP and PLDA. Because there is a large amount of features in speech recognition, considerable computational time is required to obtain an optimal projection. In order to overcome this problem, we give an efficient calculation scheme. Experimental results show that the locality preserving dimensionality reduction methods yield better performance than traditional ones.

The paper is organized as follows. Feature transformation methods are reviewed in Section 2. Existing locality preserving dimensionality reduction methods are reviewed and proposed methods are introduced in Section 3. An efficient calculation to obtain an optimal projection is given in Section 4. Experimental results are presented in Section 5. Finally, conclusions are given in Section 6.

2. LINEAR DIMENSIONALITY REDUCTION METHODS

We formulate the problem of dimensionality reduction. Given n-dimensional features $\mathbf{x}_j \in \mathbb{R}^n (j = 1, 2, \ldots, N)$, e.g., concatenated speech frames, and associated class labels $y_j \in \{1, 2, \ldots, L\}$, e.g., phonemes, let us find a projection matrix $\mathbf{B} \in \mathbb{R}^{n \times p}$ that transforms these features to p-dimensional features $\mathbf{z}_j \in \mathbb{R}^p (j = 1, 2, \ldots, N)$, where $p < n$, $\mathbf{z}_j = \mathbf{B}^T \mathbf{x}_j$, L denotes the number of classes, and N denotes the number of features. To find a projection matrix $\mathbf{B}$, we briefly review three techniques: LDA, HDA and PLDA.

2.1. Linear Discriminant Analysis

The within-class covariance matrix $\mathbf{C}^{(W)}$ and the between-class covariance matrix $\mathbf{C}^{(B)}$ are defined as follows [2, 3]:

$$\mathbf{C}^{(W)} = \frac{1}{N} \sum_{l=1}^{L} \sum_{j:y_i=l} (\mathbf{x}_j - \boldsymbol{\mu}_l)(\mathbf{x}_j - \boldsymbol{\mu}_l)^T, \tag{1}$$

$$\mathbf{C}^{(B)} = \sum_{l=1}^{L} P_l (\boldsymbol{\mu}_l - \boldsymbol{\mu})(\boldsymbol{\mu}_l - \boldsymbol{\mu})^T, \tag{2}$$

where $\boldsymbol{\mu}_l$ is the mean of features in class l, $\boldsymbol{\mu}$ is the mean of all features, and P_l is the weight of class l. LDA finds a projection matrix $\mathbf{B}$ that maximizes the following objective function:

$$J_{LDA}(\mathbf{B}) = \frac{\left|\mathbf{B}^T \mathbf{C}^{(B)} \mathbf{B}\right|}{\left|\mathbf{B}^T \mathbf{C}^{(W)} \mathbf{B}\right|}. \tag{3}$$

The following function is defined as another objective function of LDA:

$$J_{LDA}(\mathbf{B}) = tr\left((\mathbf{B}^T\mathbf{C}^{(W)}\mathbf{B})^{-1}\mathbf{B}^T\mathbf{C}^{(B)}\mathbf{B}\right). \quad (4)$$

The optimization of Eqs. (3) and (4) results in the same transformation [2].

The within-class scatter $\mathbf{S}^{(W)}$ and between-class scatter $\mathbf{S}^{(B)}$ could be employed in place of $\mathbf{C}^{(W)}$ and $\mathbf{C}^{(B)}$, respectively. The within-class scatter is defined as $\mathbf{S}^{(W)} = N\mathbf{C}^{(W)}$ and the between-class scatter is defined as $\mathbf{S}^{(B)} = N\mathbf{C}^{(B)}$. The same solution is obtained even if $\mathbf{C}^{(W)}$ and $\mathbf{C}^{(B)}$ in Eqs. (3) and (4) are replaced with $\mathbf{S}^{(W)}$ and $\mathbf{S}^{(B)}$, respectively.

2.2. Heteroscedastic Discriminant Analysis

The following objective function is defined in HDA, which considers individual weighted contributions of the class variances [4].

$$J_{HDA}(\mathbf{B}) = \prod_{l=1}^{L}\left(\frac{\left|\mathbf{B}^T\mathbf{C}^{(B)}\mathbf{B}\right|}{\left|\mathbf{B}^T\mathbf{C}_l\mathbf{B}\right|}\right)^{N_l}, \propto \frac{\left|\mathbf{B}^T\mathbf{C}^{(B)}\mathbf{B}\right|}{\prod_{l=1}^{L}\left|\mathbf{B}^T\mathbf{C}_l\mathbf{B}\right|^{P_l}}, \quad (5)$$

where N_l denotes the number of features labeled as class l and $\mathbf{C}_l$ is a class covariance matrix in class l. $\mathbf{C}_l$ is defined as

$$\mathbf{C}_l = \frac{1}{N_l}\sum_{j:y_i=l}(\mathbf{x}_j - \boldsymbol{\mu}_l)(\mathbf{x}_j - \boldsymbol{\mu}_l)^T. \quad (6)$$

The within-class covariance satisfies $\mathbf{C}^{(W)} = \sum_{l=1}^{L} P_l\mathbf{C}_l$. The solution to maximize Eq. (5) is not analytically obtained. Therefore, a numerical optimization technique is performed to maximize Eq. (5) with respect to $\mathbf{B}$.

2.3. Power Linear Discriminant Analysis

We have proposed the following objective function, which integrates LDA and HDA [5]:

$$J_{PLDA}(\mathbf{B}, m) = \frac{\left|\mathbf{B}^T\mathbf{C}^{(B)}\mathbf{B}\right|}{\left|\left(\sum_{l=1}^{L} P_l(\mathbf{B}^T\mathbf{C}_l\mathbf{B})^m\right)^{1/m}\right|}, \quad (7)$$

where m denotes a control parameter. We have referred to it as power linear discriminant analysis (PLDA). Intuitively, as m becomes larger, the classes with larger variances become dominant in the denominator of Eq. (7). Conversely, as m becomes smaller, the classes with smaller variances become dominant. Thus, by varying the control parameter m, the objective function can represent various objective functions. If m is set to one/zero, the objective function agrees with the LDA/HDA objective function [8]. One issue regarding PLDA in practice is how to select the optimal control parameter m. In [9], the selection method of an optimal control parameter is provided.

3. DIMENSIONALITY REDUCTION PRESERVING LOCALITY OF DATA STRUCTURE

Recently, several dimensionality reduction methods for multimodal data have been proposed in the machine learning community. We review locality preserving projection (LPP) and local Fisher discriminant analysis (LFDA). Then, we propose locality preserving HDA, which combines the ideas of LPP and HDA, and locality preserving PLDA, which combines the ideas of LPP and PLDA.

3.1. Locality Preserving Projection

Let $\mathbf{A}$ be a symmetric $n \times n$ matrix, which is called an affinity matrix. The (i,j)-element A_{ij} of $\mathbf{A}$ is the affinity between $\mathbf{x}_i$ and $\mathbf{x}_j$. The affinity A_{ij} becomes a large value if $\mathbf{x}_i$ and $\mathbf{x}_j$ are located in the near distance. Contrarily, A_{ij} becomes a small value if $\mathbf{x}_i$ and $\mathbf{x}_j$ are located in the far distance. There are several different definitions of $\mathbf{A}$, e.g., a heat kernel or a nearest neighbor. The objective function of LPP is defined as follows [6]:

$$J_{LPP}(\mathbf{B}) = \frac{1}{2}\sum_{i,j=1}^{N} A_{ij}||\mathbf{B}^T\mathbf{x}_i - \mathbf{B}^T\mathbf{x}_j||^2, \quad (8)$$

$$\text{s. t. } \mathbf{B}^T\mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{B} = \mathbf{I}, \quad (9)$$

where $\mathbf{X} = [\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_N]$, $\mathbf{I}$ is the identity matrix, and $\mathbf{D}$ is a diagonal matrix whose (i,i)-element is as follows: $D_{i,i} = \sum_{j=1}^{N} A_{i,j}$.

In LPP, the data pairs close to each other in the original space remain close in the projected space. LPP is an unsupervised dimensionality reduction method.

3.2. Local Fisher Discriminant Analysis

Local Fisher discriminant analysis (LFDA) [7] has been proposed by Sugiyama, which combines the ideas of LDA (FDA) and LPP. He reformulated the within-class scatter $\mathbf{S}^{(W)}$ and the between-class scatter $\mathbf{S}^{(B)}$ in a pairwise manner:

$$\mathbf{S}^{(W)} = \frac{1}{2}\sum_{i,j=1}^{N} W_{ij}^{(W)}(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T, \quad (10)$$

$$\mathbf{S}^{(B)} = \frac{1}{2}\sum_{i,j=1}^{N} W_{ij}^{(B)}(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T, \quad (11)$$

where

$$W_{ij}^{(W)} = \begin{cases} 1/N_l & \text{if } y_i = y_j = l, \\ 0 & \text{if } y_i \neq y_j, \end{cases} \quad (12)$$

$$W_{ij}^{(B)} = \begin{cases} 1/N - 1/N_l & \text{if } y_i = y_j = l, \\ 1/N & \text{if } y_i \neq y_j. \end{cases} \quad (13)$$

Based on an affinity matrix $\mathbf{A}$ and the pairwise expressions of the between/within class scatter, a *local* within-class scatter and a *local* between-class scatter are defined as follows [7]:

$$\tilde{\mathbf{S}}^{(W)} = \frac{1}{2}\sum_{i,j=1}^{N} \tilde{W}_{ij}^{(W)}(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T, \quad (14)$$

$$\tilde{\mathbf{S}}^{(B)} = \frac{1}{2}\sum_{i,j=1}^{N} \tilde{W}_{ij}^{(B)}(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T, \quad (15)$$

where

$$\tilde{W}_{ij}^{(W)} = \begin{cases} A_{ij}/N_l & \text{if } y_i = y_j = l, \\ 0 & \text{if } y_i \neq y_j, \end{cases} \tag{16}$$

$$\tilde{W}_{ij}^{(B)} = \begin{cases} A_{ij}(1/N - 1/N_l) & \text{if } y_i = y_j = l, \\ 1/N & \text{if } y_i \neq y_j. \end{cases} \tag{17}$$

Both $\tilde{\mathbf{S}}^{(W)}$ and $\tilde{\mathbf{S}}^{(B)}$ put a weight on sample pairs in the same class. The objective function of LFDA is defined as follows:

$$J_{LFDA}(\mathbf{B}) = tr\left(\left(\mathbf{B}^T\tilde{\mathbf{S}}^{(W)}\mathbf{B}\right)^{-1}\mathbf{B}^T\tilde{\mathbf{S}}^{(B)}\mathbf{B}\right). \tag{18}$$

LFDA searches for a projection matrix $\mathbf{B}$ such that nearby data pairs in the same class remain close and the data pairs in different classes are separated from each other; far-apart pairs in the same class are not forced to be close. Thus, LFDA is a supervised dimensionality reduction method. If A_{ij} is taken to be one for all in-class pairs, LFDA corresponds exactly to LDA because $\tilde{\mathbf{S}}^{(W)}$ and $\tilde{\mathbf{S}}^{(B)}$ agree with $\mathbf{S}^{(W)}$ and $\mathbf{S}^{(B)}$, respectively.

In the same fashion as the definition of LDA objective functions, the following function is defined as another objective function of LFDA:

$$J_{LFDA}(\mathbf{B}) = \frac{\left|\mathbf{B}^T\tilde{\mathbf{S}}^{(B)}\mathbf{B}\right|}{\left|\mathbf{B}^T\tilde{\mathbf{S}}^{(W)}\mathbf{B}\right|}. \tag{19}$$

The optimization of Eqs. (18) and (19) results in the same projection.

The local within-class covariance $\tilde{\mathbf{C}}^{(W)}$ and the local between-class covariance $\tilde{\mathbf{C}}^{(B)}$ can be defined as $\tilde{\mathbf{C}}^{(W)} = \frac{1}{N}\tilde{\mathbf{S}}^{(W)}$ and $\tilde{\mathbf{C}}^{(B)} = \frac{1}{N}\tilde{\mathbf{S}}^{(B)}$, respectively. The same solution is obtained when $\tilde{\mathbf{S}}^{(W)}$ and $\tilde{\mathbf{S}}^{(B)}$ in Eqs. (18) and (19) are replaced with $\tilde{\mathbf{C}}^{(W)}$ and $\tilde{\mathbf{C}}^{(B)}$, respectively.

3.3. Local Heteroscedastic Discriminant Analysis

Inspired by LFDA, we combine the ideas of LPP and HDA. Let us define a *local* class covariance matrix $\tilde{\mathbf{C}}_l$ as follows:

$$\tilde{\mathbf{C}}_l = \frac{1}{2N_l}\sum_{i,j=1}^{N}\tilde{W}_{ij}^{l}(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T, \tag{20}$$

where

$$\tilde{W}_{ij}^{l} = \begin{cases} A_{ij}/N_l & \text{if } y_i = y_j = l, \\ 0 & \text{otherwise.} \end{cases} \tag{21}$$

From Eqs. (16) and (21), $\tilde{W}_{ij}^{(W)} = \sum_{l=1}^{L}\tilde{W}_{ij}^{l}$. In addition, $\tilde{\mathbf{C}}^{(W)}$ satisfies $\tilde{\mathbf{C}}^{(W)} = \sum_{l=1}^{L}P_l\tilde{\mathbf{C}}_l$. We define the following objective function:

$$J_{LHDA}(\mathbf{B}) = \frac{\left|\mathbf{B}^T\tilde{\mathbf{C}}^{(B)}\mathbf{B}\right|}{\prod_{l=1}^{L}\left|\mathbf{B}^T\tilde{\mathbf{C}}_l\mathbf{B}\right|^{P_l}}, \tag{22}$$

which is called local HDA (LHDA). If A_{ij} is taken to be one for all in-class pairs, LHDA corresponds exactly to HDA because $\tilde{\mathbf{C}}_l$ agrees with $\mathbf{C}_l$.

3.4. Local Power Linear Discriminant Analysis

As in the case of LHDA, using *local* class covariances $\tilde{\mathbf{C}}_l$, we extend the PLDA objective function as follows:

$$J_{LPLDA}(\mathbf{B}, m) = \frac{\left|\mathbf{B}^T\tilde{\mathbf{C}}^{(B)}\mathbf{B}\right|}{\left|\left(\sum_{l=1}^{L}P_l(\mathbf{B}^T\tilde{\mathbf{C}}_l\mathbf{B})^m\right)^{1/m}\right|}, \tag{23}$$

called local PLDA (LPLDA). From Eqs. (19) and (22), LPLDA with m=1 agrees with LFDA, and LPLDA with m=0 agrees with LHDA. LPLDA corresponds exactly to PLDA when A_{ij} is taken to be one for all in-class pairs.

4. EFFICIENT COMPUTATION OF LOCAL CLASS COVARIANCES AND LOCAL BETWEEN-CLASS COVARIANCE

To obtain optimal projections of the LHDA and LPLDA objective functions, the calculations of $\tilde{\mathbf{C}}_l$ and $\tilde{\mathbf{C}}^{(B)}$ are needed in advance. Both matrices require N^2 times summation. Because acoustic models in a speech recognition system are generally trained using a large amount of speech data, the value of N tends to become large , e.g., 10^6 to 10^8. Therefore, the computational costs of $\tilde{\mathbf{C}}_l$ and $\tilde{\mathbf{C}}^{(B)}$ tend to be huge.

In order to calculate $\tilde{\mathbf{C}}_l$ and $\tilde{\mathbf{C}}^{(B)}$ efficiently, we assume that a distribution of each class is constructed from several clusters, that $\mathbf{x}_i (i = 1, \ldots, N)$ is generated from one of the clusters, and that the number of clusters in each class is set in advance. We can redefine a local class covariance as follows:

$$\tilde{\mathbf{C}}_l = \sum_{m=1}^{M_l}P_{l,m}\mathbf{C}_{l,m}, \tag{24}$$

where M_l is the number of clusters in class l, $P_{l,m}$ is the weight of the m-th cluster in class l, and $\mathbf{C}_{l,m}$ denotes an m-th cluster covariance in class l. $\tilde{\mathbf{C}}_l$ in Eq. (20) is equal to $\tilde{\mathbf{C}}_l$ in Eq. (24) when the affinity matrix is defined as follows: $A_{ij} = 1/P_{l,m}$ if $\mathbf{x}_i$ and $\mathbf{x}_j$ belong to the same cluster m in a class l, otherwise $A_{ij} = 0$. To obtain $P_{l,m}$ and $\mathbf{C}_{l,m}$, we employ the Expectation-Maximization (EM) algorithm. We can efficiently calculate $\tilde{\mathbf{C}}_l$ to employ this computational scheme under the above assumptions because N^2 times summation is no longer required in Eq. (24).

Under the same assumptions, local between-class covariance can be rewritten as

$$\tilde{\mathbf{C}}^{(B)} = \mathbf{C}^{(T)} - \sum_{l=1}^{L}P_l\left(P_l\mathbf{C}_l + (1 - P_l)\sum_{m=1}^{M_l}P_{l,m}\mathbf{C}_{l,m}\right), \tag{25}$$

where $\mathbf{C}^{(T)}$ denotes a total covariance matrix, which is defined as $\mathbf{C}^{(T)} = \mathbf{C}^{(B)} + \mathbf{C}^{(W)}$. Once we calculate $\mathbf{C}^{(T)}$ and $\mathbf{C}_l$, and estimate $P_{l,m}$ and $\mathbf{C}_{l,m}$ using the EM algorithm, we can calculate $\tilde{\mathbf{C}}^{(B)}$ immediately.

5. EXPERIMENTS

We conducted experiments using the CENSREC-3 database [10]. CENSREC-3 is designed as an evaluation framework of Japanese isolated word recognition in real car-driving environments. Speech data were collected using 2 microphones: a close-talking (CT) microphone and a hands-free (HF) microphone. For training, a driver's

speech of phonetically-balanced sentences was recorded under two conditions: while idling and while driving on a city street with a normal in-car environment. A total of 28,100 utterances spoken by 293 drivers (202 males and 91 females) were recorded with both microphones. For evaluation, a driver's speech of isolated words was recorded under 16 environmental conditions using combinations of three kinds of vehicle speeds (idling, low-speed driving on a city street, and high-speed driving on an expressway) and six kinds of in-car environments (normal, with hazard lights on, with the air-conditioner on (fan low/high), with the audio CD player on, and with windows open). In these conditions, the "hazard lights on" condition was used only when idling. We used only 14,050 utterances recorded with a CT microphone for training. We only used three kinds of vehicle speeds in the normal in-car environment for evaluation. A total of 2,646 utterances spoken by 18 speakers (8 males and 10 females) and collected using a CT microphone were evaluated. The speech signals for training and evaluation were both sampled at 16 kHz. The decoding process is performed without any language model. The vocabulary size is 100 words, which includes the original fifty words and another fifty similar-sounding words.

5.1. Baseline System

In CENSREC-3, the baseline scripts are designed to facilitate HMM training and evaluation by HTK [11]. The acoustic models consist of triphone HMMs. Each HMM has five states and three of them had output distributions. Each distribution is represented with 32 mixture diagonal Gaussians. The total number of states with distributions is 2,000. The feature vector consists of 12 MFCCs and log-energy with their corresponding delta and acceleration coefficients (total 39 dimensions). Frame length is 20 ms and frame shift is 10 ms. In the Mel-filter bank analysis, a cut-off is applied to frequency components lower than 250 Hz.

5.2. Feature Transformation Procedure

Feature transformation was performed using LDA+MLLT [12], HDA+MLLT [4], PLDA [5], LFDA [7] +MLLT , LHDA+MLLT, and LPLDA for the spliced features. Eleven successive frames (143 dimensions) were reduced to 39 dimensions. In PLDA and LPLDA, we assumed that projected class covariance matrices were diagonal and used the limited-memory BFGS algorithm as a numerical optimization technique, and their control parameters were experimentally selected. The LDA transformation matrix was used as the initial gradient. In LFDA, LHDA and LPLDA, the number of mixtures was four for each class, while the number of mixtures was one for the classes that have training data of less than one percent of the total. In addition, to obtain an optimal projection matrix, we employed an efficient computation scheme for calculating covariances. To assign one of the classes to every feature vector, HMM state labels were generated for the training data by a state-level forced alignment algorithm using a well-trained HMM system. The number of classes was 40, corresponding to the number of the monophones.

5.3. Results

We performed the experiments using the above feature transformations on CENSREC-3. The results are presented in Table 1. Among dimensionality reduction methods for unimodal data, i.e. LDA+MLLT, HDA+MLLT and PLDA, the lowest WER was obtained by PLDA with $m = -0.5$. Among all dimensionality reduction methods, LPLDA with $m = -0.25$ yielded the lowest WER. The locality preserving dimensionality reduction methods consistently yielded better performance than the traditional methods.

Table 1. Word error rates (%) using feature transformation methods. The best results are highlighted in bold.

	WER		WER
MFCC + Δ +$\Delta\Delta$	6.50		
LDA+MLLT	6.12	LFDA+MLLT	5.10
HDA+MLLT	7.14	LHDA+MLLT	6.43
PLDA (m =-0.25)	6.50	LPLDA (m =-0.25)	**5.03**
PLDA (m =-0.5)	5.67	LPLDA (m =-0.5)	5.40

6. CONCLUSIONS

In this paper we introduced two dimensionality reduction methods, HDA preserving the local structure of the data and PLDA preserving the local structure. Experimental results showed that locality preserving methods yielded better performance than traditional methods. The best performance was obtained by LPLDA, and it was about 22% relatively better than the performance of the baseline system.

7. ACKNOWLEDGMENT

The study presented was conducted using the CENSREC-3 database developed by the IPSJ-SIG SLP Noisy Speech Recognition Evaluation Working Group.

8. REFERENCES

[1] S. Nakagawa and K. Yamamoto, "Evaluation of segmental unit input HMM," *Proc. ICASSP*, pp. 439–442, 1996.

[2] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, New York, second edition, 1990.

[3] R. O. Duda, P. B. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, New York, 2001.

[4] G. Saon, M. Padmanabhan, R. Gopinath, and S. Chen, "Maximum likelihood discriminant feature spaces," *Proc. ICASSP*, pp. 129–132, 2000.

[5] M. Sakai, N. Kitaoka, and S. Nakagawa, "Generalization of linear discriminant analysis used in segmental unit input HMM for speech recognition," *Proc. ICASSP*, pp. 333–336, 2007.

[6] X. He and P. Niyogi, "Locality preserving projections," in *Advances in Neural Information Processing Systems*, 2004.

[7] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis," *Journal of Machine Learning Research*, vol. 8, pp. 1027–1061, 2007.

[8] M. Sakai, N. Kitaoka, and S. Nakagawa, "Linear discriminant analysis using a generalized mean of class covariances and its application to speech recognition," *IEICE Transactions on Information and Systems*, vol. E91-D, no. 3, pp. 478–487, 2008.

[9] M. Sakai, N. Kitaoka, and S. Nakagawa, "Selection of optimal dimensionality reduction method using Chernoff bound for segmental unit input HMM," *Proc. Interspeech*, pp. 1110–1113, 2007.

[10] M. Fujimoto, K. Takeda, and S. Nakamura, "CENSREC-3: An evaluation framework for Japanese speech recognition in real driving-car environments," *IEICE Transactions on Information and Systems*, vol. E89-D, no. 11, pp. 2783–2793, 2006.

[11] *HTK Web site*, `http://htk.eng.cam.ac.uk/`.

[12] R. A. Gopinath, "Maximum likelihood modeling with Gaussian distributions for classification," *Proc. ICASSP*, 1998.