ABSTRACT
Linear Discriminant Analysis (LDA) is a well-known technique to improve the discrimination among classes or reduce the dimensionality with minimum loss of discrimination. It is generally used as part of the front-end of a speech recognizer with the classes defined on phone or subphone level. LDA with subphone-level classes (such as tied states or individual Gaussian densities) shows superior performance to that with phone-level classes. However, it does not seem to provide the best discrimination from viewpoint of recognition performance.

This paper focuses on improving the discrimination between phones while subphones such as tied states are used as the basic classes of LDA. To this end, this paper proposes a method to control the discrimination ability among subphone classes so as to de-emphasize (or ignore) the discrimination between those belonging to the same phone, while emphasizing the discrimination between those belonging to the different phones.

Proposed method is implemented in the framework of Weighted Pairwise Scatter LDA (WPS-LDA) and is evaluated on Korean connected digit recognition task. According to experimental results, proposed method shows higher performance than conventional LDA and WPS-LDA. Performance is the highest when the discrimination between pairwise classes belonging to the same phone is totally ignored. The string error reduction rates of conventional LDA, WPS-LDA and proposed method are 13%, 14.5% and 23%, respectively.

1. INTRODUCTION
Linear Discriminant Analysis (LDA) is a widely used technique to improve the discrimination among classes in pattern classification fields. In speech recognition, the successful usage of it at the front-end reduces dimensions while minimizing the loss of discrimination or improves the recognition accuracy. However, LDA has not always improved the performance of speech recognition systems mainly because it is based on some implicit assumptions not satisfied in real situations. To deal with this problem, various methods have been developed including Weighted Pairwise Scatter Linear Discriminant Analysis (WPS-LDA) [5], Heteroscedastic Discriminant Analysis (HDA) [6][7], Mutual Information Discriminant Analysis (MIDA) [8] and linear transformations based on Minimum Classification Error (MCE) [10].

Li et al. put weights on between-class scatter matrix based on the distance between pairwise classes to remove the LDA assumption that each class is equally confusable with all other classes [5]. Kumar et al. generalized LDA to deal with the cases of different within-class covariances by dropping the assumption of LDA that each class has equal within-class covariance [6]. Saon et al. introduced the criteria based on the minimum Bayes’s error [9]. However methods except [5] could not discover the closed form solution to find transformation matrix. Instead they used the gradient descent method to find the optimal transformation matrix. The approach in [5] is simpler than other methods [6]-[10] because it employs conventional LDA method except differently obtaining the between-class scatter matrix.

As mentioned in [3], it is not clear what the most appropriate class definition for LDA should be. It has been reported that the most recognition systems show the best performance when using subphone-level classes (such as the state or tied states and individual Gaussian densities) instead of phone-level classes [3]-[10]. However, LDA with subphone-level classes does not seem to provide the best discrimination because it considers pairwise classes belonging to the same phone and those belonging to different phones equally. In order to improve phone-level discrimination with subphone classes, this paper proposes a method to control the discrimination ability among subphone classes so as to de-emphasize (or ignore) the discrimination between those belonging to the same phone, while emphasizing the discrimination between those belonging to the acoustically similar phones. In this paper, proposed method is implemented in the framework of WPS-LDA.

In section 2 we briefly describe the conventional LDA and WPS-LDA and also problem they possess. In section 3 we introduce new weighting method between classes of subphone level in order to improve the discrimination among phones. In section 4 we discuss various experiments and the results of conventional methods and the new method. Finally, section 5 provides the conclusion.

2. LDA
2.1. Conventional LDA
The LDA with Fisher’s criterion is formulated as follows. Let \( x \in \mathbb{R}^n \) be a feature vector. Total number of feature vectors is \( N \) and each \( x \) has a label of the class they belonged to. We seek to find a transformation matrix \( A \) with minimum loss of discrimination for classification.
\[ y = A^T x \quad A : \mathbb{R}^p \rightarrow \mathbb{R}^n \] (1)

where \( p < n \). We first define within-class scatter matrix and between-class scatter matrix. Within-class scatter matrix is defined by

\[ S_w = \sum_{i=1}^{L} N_i \Sigma_i \] (2)

where \( L \) is the number of classes, \( N_i \) is the number of feature vectors in class \( i \), and \( \Sigma_i \) is the sample covariance matrix of class \( i \). Between-class scatter matrix is defined by

\[ S_b = \sum_{i=1}^{L} N_i (\mu_i - \mu)(\mu_i - \mu)^T \] (3)

where \( \mu_i \) is the sample mean of class \( i \) and \( \mu \) is the mean of total samples. It is straightforward that \( S_b \) can be represented as

\[ S_b = \frac{1}{2N} \sum_{i=1}^{L} \sum_{j=1}^{L} N_i N_j (\mu_i - \mu)(\mu_i - \mu)^T \] (4)

Among the several criteria used for LDA, we employed

\[ J = \text{tr}(S_b^{-1} S_w) \] (5)

Then \( A \) is the transformation matrix to maximize (5). As well known, we can easily find \( A \) from

\[ S_{\Phi} = S_b^{-1} S_w \Phi = \Phi A \] (6)

where \( \Phi \) and \( A \) are \( n \times n \) matrix of eigenvectors and eigenvalues of \( S_{\Phi} \), respectively. \( A \) is made by choosing \( p \) eigenvectors corresponding to \( p \) largest eigenvalues.

### 2.2. WPS-LDA [5]

One of implicit assumptions in LDA is that each class is equally confusable with all other classes. It causes the problem that the between-class scatter matrix mostly ignores the discriminatory information between the classes that are close to each other.

To alleviate this problem, Li et al. proposed WPS-LDA which imposes weighting factor to class when calculating between-class scatter matrix so that more confusable classes may be weighted more than less confusable classes. To impose the weight on \( S_b \), (3) or (4) is modified as follows:

\[ S_{b,\text{WPS}} = \frac{1}{2N} \sum_{i=1}^{L} \sum_{j=1}^{L} N_i N_j w_{ij} (\mu_i - \mu)(\mu_i - \mu)^T \] (7)

where \( w_{ij} \) is a non-negative weight assigned to class pair \((i, j)\). \( S_{b,\text{WPS}} \) is equivalent to \( S_b \) when all \( w_{ij} = 1 \). In WPS-LDA, \( S_{b,\text{WPS}} \) is used to find \( A \). Among a few choices for \( w_{ij} \) as a function of either Euclidean or Kullback-Leibler distance, the best performance was obtained when the weight is equal to the square of the inverse of the Euclidean distance between class means.

\[ w_{ij} = \frac{1}{(\mu_i - \mu_j)^T (\mu_i - \mu_j)} = \frac{1}{d_{ij}^2} \] (8)

In WPS-LDA, the closer the distance between pairwise classes is, the larger weight value is. Therefore, WPS-LDA can obtain the better discrimination among defined classes than conventional LDA. However, if classes in WPS-LDA are defined at subphone level, since distance between classes with same phone is usually smaller than that with different phones, weight value between pairwise classes belonging to same phone may be increased too much. This unwanted problem may result in performance degradation.

### 3. NEW WEIGHTING METHOD BASED ON WPS-LDA

It is important to maximize the phone-level discrimination from the viewpoint of recognition performance. However it seems to be difficult to accomplish high discrimination on that level. Fig.1 illustrates an example of this problem. Let’s say that the acoustic spaces of two confusable phones ‘A’ and ‘B’ are represented by dotted lines in Fig.1(a). If modeled with single Gaussian density, they might be represented by dotted lines in Fig.1(b) and noticeable portions of them are overlapped. In this case the distribution can be modeled by multi-mixture Gaussian densities as shown by solid lines in Fig.1(a) and (b). From this example, it can be easily understood that LDA with subphone-level classes such as tied states and individual Gaussian densities shows better performance than that with phone-level classes.

![Figure 1: An example representing distributions of two phones (a) acoustic spaces (dotted line) of phone ‘A’ and ‘B’ and multi-mixture Gaussian modeling (solid line) (b) Gaussian modeling of phones (dotted line - single mixture) or subphone level (solid line - multi-mixture)](image)

However, improving discrimination on the subphone level does not necessarily increase speech recognition performance,
because it considers discrimination between pairwise subphone classes belonging to the same phone and that belonging to different phones equally. To improve phone-level discrimination in LDA with subphone classes, it is necessary to control the discrimination ability among subphone classes so as to de-emphasize (or ignore) the discrimination between those belonging to the same phone, while emphasizing the discrimination between those belonging to the different phones. This approach is especially important when WPS-LDA is employed. As mentioned in Section 2.2, WPS-LDA with subphone classes tends to irrelevantly emphasize discrimination between acoustically similar classes that belong to the same phone.

Based on this concept, \( S_{b,WPS} \) in (7) can be modified as

\[
S_{b,WPS} = \frac{1}{2N} \sum_{i} \left( S_{b,same}^{i} + S_{b,diff}^{i} \right)
\]  

(9)

where

\[
S_{b,same}^{i} = \sum_{j \in \Omega_{i}} N \cdot N_{ij} \cdot d_{ij}^{2} \cdot w_{ij}^{same} (\mu_{ij} - \mu_{j})' (\mu_{ij} - \mu_{j})
\]

(10)

\[
S_{b,diff}^{i} = \sum_{j \in \Omega_{i}} N \cdot N_{ij} \cdot d_{ij}^{2} \cdot w_{ij}^{diff} (\mu_{ij} - \mu_{j})' (\mu_{ij} - \mu_{j})
\]

(11)

and \( \Omega_{i} \) is a set of class index such that if \( j \in \Omega_{i} \), then class \( j \) belongs to the same base phone as class \( i \) does and vice versa. \( w_{ij}^{same} \) and \( w_{ij}^{diff} \) are the weight between classes belonging to the same phone and that belonging to the different phones, respectively. We define \( w_{ij}^{same} \) and \( w_{ij}^{diff} \) by

\[
w_{ij}^{same} = f_{same}(d_{ij}^{2}), \quad w_{ij}^{diff} = f_{diff}(d_{ij}^{2})
\]

(12)

Here an important issue is how to determine \( f_{same}(\cdot) \) and \( f_{diff}(\cdot) \) to improve the discrimination on phone level. In this paper, among many possibilities, we chose

\[
f_{same}(d_{ij}^{2}) = C, \quad f_{diff}(d_{ij}^{2}) = \frac{1}{d_{ij}^{2}}
\]

(13)

where \( C \) is a constant which is to be determined experimentally such that \( w_{ij}^{same} \ll w_{ij}^{diff} \) in almost all cases. Among the choice of \( C \), \( C = 0 \) or \( w_{ij}^{same} = 0 \) is a special case that contribution of pairwise between-class scatter belonging to the same phone is totally ignored.

4. EXPERIMENTS AND RESULTS

4.1. Experiments

We evaluated the proposed method with speaker-independent Korean connected digit recognition experiments. All Korean digits are monosyllables and many of them are acoustically very confusable.

Speech database was gathered over wired and wireless telephone network and was sampled at 8kHz. 5562 utterances spoken by 175 male speakers were used for training, and 2513 utterances spoken by remaining 80 speakers were used for test.

We used 38 dimensional feature parameters (12 MFCCs, its deltas and double deltas, delta log energy and double delta log energy). And we employed the cepstral mean subtraction (CMS) to compensate for telephone channel distortion.

Our baseline system used triphones with continuous mixture density HMM. Each HMM has five states and the number of mixtures per state varied from 1 to 9. We used state tying based on tree based clustering (TBC) method. Total number of tied states was 447 and we used these tied states as basic classes for LDA.

4.2. Results and Discussions

The performances of baseline system, conventional LDA and WPS-LDA are shown in Fig.2. The performance of WPS-LDA based on Euclidean distance is worse than conventional LDA. This result may be caused by problem of WPS-LDA described in section 2.2. In WPS-LDA, weight becomes larger without limitation when the pairwise classes are closer. If weight value range is constrained, unwanted effect of large weight value for the pairwise classes belonging to the same phone can be reduced. In this paper, to normalize the dynamic range of weight, we tried a new type of weight,

\[
w_{ij} = 11 \cdot \text{sigmoid}(d_{ij}^{2}) - T
\]

(14)

where \( \text{sigmoid}(x) = \frac{1}{1+e^{-\gamma x}} \), \( \gamma \) is the slope of sigmoid function and \( T \) is a constant. The dynamic range of the weight in (14) is between 0 and 1 when \( T = 1 \). As shown in Fig.2, performance of WPS-LDA based on sigmoid function is higher than that of WPS-LDA using weight in (8), which confirms our assertion in Section 2.2 and Section 3.

![Figure 2: Comparison of results of several methods. WPS(Euc) and WPS(sig) indicate WPS-LDA using weight based on Euclidean distance as in (8) and that based on sigmoid function as in (14), respectively.](image-url)
In Fig.3, we also compared the proposed method with the methods mentioned above. The proposed method showed the highest performance. The string error reduction rates of conventional LDA, WPS-LDA and proposed method are 13%, 14.5% and 23%, respectively. In this experiment, we used $C = 0$ or $w^{new}_i = 0$ to ignore the contribution of pairwise scatter between classes belonging to the same phone.

To explore the effect of $w^{new}_i$ or $C$ in (13), we experimented with different values of $C$. The result can be seen in Fig.4. As the figure shows, setting $C = 0$ yields the best performance. It means that removing the influence among classes belonging to the same phone is the most effective.

We also conducted the experiment for the proposed method using weights based on sigmoid function as in (14). Although performance was higher than conventional LDA but was lower than that based on Euclidean distance. It is because weight method based on Euclidean distance imposes larger weight than that based on the sigmoid function when two classes with different phones is confusable although discrimination between classes with same phone is ignored.

5. CONCLUSIONS

A new weighting scheme is introduced into LDA with subphone-level classes such as tied states or individual Gaussian densities to improve the discrimination of phone level. Different weighting factors were applied to pairwise scatter depending on whether the pairwise subphone classes belong to the same phone or not. Experimental results show that the proposed method outperforms both the conventional LDA and WPS-LDA. Best performance was achieved when contribution of pairwise scatter belonging to the same phone is ignored. Compared to the baseline system, the string error reduction rate by the proposed method was 23%.

Performance evaluation with dimensionality reduction by proposed method is now under way. Our future work includes evaluation on the large vocabulary recognition task domain.

6. REFERENCES